

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 5, 2003, 22:00:00 ; Search time 1652 Seconds
(without alignments)
2360.639 Million cell updates/sec

Title: US-10-048-196-2

Perfect score: 676

Sequence: 1 MKKLYVTAILMTAFTLASC.....SLKPCMLETVNAFIVPTTTR 134

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US10048196/runat_28042003_151439_5581/app_query.fasta_1.327
-DB=GenEmbl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10048196 -CGN_1_1687_@runat_28042003_151439_5581 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_ov.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_cm.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	676	100.0	405	6	AX081157 Sequence
2	676	100.0	39003	6	AX067446 Sequence
3	118	17.5	329861	1	NMA522491
C 4	116	17.2	11489	1	AE002455 Neisseria
C 5	116	17.2	349980	6	AX044031 Sequence
6	103.5	15.3	68466	2	AP000686 Homo sapi
7	103.5	15.3	187087	2	AC079123 Homo sapi
8	103.5	15.3	188041	9	AP003327 Homo sapi
9	103.5	15.3	238472	9	AC010290 Homo sapi
10	97	14.3	5164	9	AB058719 Homo sapi
11	94.5	14.0	780	8	BTAGMP
12	94	13.9	249846	2	AC125227 Mus muscu
13	92.5	13.7	632	9	AF223572 Homo sapi
C 14	92.5	13.7	174410	9	AC097464 Homo sapi
C 15	92	13.6	226325	2	AL845171 Mus muscu
C 16	90.5	13.4	62043	9	AL161893 Human DNA
17	89.5	13.2	1102	8	WMTGLIABA
C 18	89	13.2	27081	3	LMFL163
C 19	89	13.2	184024	2	AC127124 Rattus no
C 20	88.5	13.1	5857	6	BD003708 Polynucle
21	88.5	13.1	11849	1	AE007414 Streptoco
22	88.5	13.1	227634	2	AC111203 Rattus no
23	88	13.0	2000	6	I03335 Sequence 2
24	88	13.0	2347	8	TAGLIAA
C 25	88	13.0	157362	2	CNS08CAU
26	87	12.9	810	8	TAE133611
27	87	12.9	1081	8	TSP130948
28	87	12.9	1672	8	TAGLIAAG2
29	87	12.9	3022	8	TAGLIA
30	87	12.9	3022	8	TAE13307
31	87	12.9	3022	8	WMTGLIABE
C 32	87	12.9	149145	8	AP003251 Oryza sat
33	86.5	12.8	3034	8	TAE13306
34	86.5	12.8	133255	2	AC121410 Rattus no
C 35	86.5	12.8	185108	2	AC118835 Rattus no
36	86	12.7	822	8	TAE133604
C 37	85.5	12.6	3823	6	AR204236
C 38	85.5	12.6	11010	1	AE005438 Escherich
C 39	85.5	12.6	157377	2	AC123926 Mus muscu
40	85.5	12.6	243098	2	AC124403 Mus muscu
C 41	85.5	12.6	268857	1	AP002560 Escherich
42	85	12.6	2976	1	AF226283 Xanthomon
43	84.5	12.5	834	8	TAE133610
44	84.5	12.5	3676	14	AB064602 TT virus
C 45	84.5	12.5	172372	3	AC010110 Drosophil

ALIGNMENTS

RESULT 1

AX081157
 LOCUS AX081157 405 bp DNA linear PAT 27-FEB-2001
 DEFINITION Sequence 1 from Patent WO0109331.
 ACCESSION AX081157
 VERSION AX081157.1 GI:13170049
 KEYWORDS
 SOURCE Moraxella catarrhalis.
 ORGANISM Moraxella catarrhalis
 Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 Moraxella.
 REFERENCE 1 (bases 1 to 405)
 AUTHORS Thonard,J.S.
 TITLE Moraxella catarrhalis antigen basb125
 JOURNAL Patent: WO 0109331-A 1 08-FEB-2001;
 SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
 FEATURES
 source Location/Qualifiers
 1..405
 /organism="Moraxella catarrhalis"
 /db_xref="taxon:480"
 BASE COUNT 130 a 108 c 72 g 95 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5 34e-62 Length: 405
 Score: 676.00 Matches: 134
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-048-196-2 (1-134) x AX081157 (1-405)
 Qy 1 MetMetLysLeuLeuValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
 Db 1 ATGATGAAATTTTATATGTTACCGCCACACTAATGACTGCCTTTACCTTAGCAAGCTGT 60
 Qy 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
 Db 61 GCAAGTACCCCTGAAGCAATCCAAAAAACAGTTCTGCTAATTTAACCCAGCTTAATC 120
 Qy 41 LysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIle 60
 Db 121 AAACATGCAGTCAACCAACTGTCAAAACCCAGCTGCTGCTAATTTAACCCAGCTTAATC 180
 Qy 61 AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCys 80
 Db 181 GCCGCCATGAATTTGCTTCAGAAATCCAAAGCCAAAGATTTTCAGAAACGGCATCGGCTGT 240
 Qy 81 ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
 Db 241 GTGGCTGATAAAGCACCTGAAGCGCTGAGCTTGACCGAGCTTACCAGAGCTGCCATTAAT 300
 Qy 101 ProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMet 120
 Db 301 CCAATGCACGCACTGAAGTTGCCAAAAAATTTGCGACACTCGCTTAAACCTTGCATG 360
 Qy 121 LeuGluThrValAsnAlaPheIleValProThrThrArg 134
 Db 361 CTAGAGACCGTCATGCTTTTATTGTGCCAACTACCACAGC 402
 RESULT 2
 AX067446
 LOCUS AX067446 39003 bp DNA linear PAT 24-JAN-2001
 DEFINITION Sequence 21 from Patent WO0078968.
 ACCESSION AX067446
 VERSION AX067446.1 GI:12545066
 KEYWORDS
 SOURCE Moraxella catarrhalis.
 ORGANISM Moraxella catarrhalis
 Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 Moraxella.
 REFERENCE 1 (bases 1 to 39003)
 AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.

TITLE Nucleotide sequences of moraxella catarrhalis genome
 JOURNAL Patent: WO 0078968-A 21 28-DEC-2000;
 Incyte Genomics, Inc. (US)
 FEATURES
 source Location/Qualifiers
 1..39003
 /organism="Moraxella catarrhalis"
 /db_xref="taxon:480"
 BASE COUNT 11568 a 8751 c 7476 g 11208 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.58e-59 Length: 39003
 Score: 676.00 Matches: 134
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-048-196-2 (1-134) x AX067446 (1-39003)
 Qy 1 MetMetLysLeuLeuValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
 Db 6901 ATGATGAAATTTTATATGTTACCGCCACACTAATGACTGCCTTTACCTTAGCAAGCTGT 6900
 Qy 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
 Db 6961 GCAAGTACCCCTGAAGCAATCCAAAAAACAGTTCTGCTAATTTAACCCAGCTTAATC 7020
 Qy 41 LysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIle 60
 Db 7021 AAACATGCAGTCAACCAACTGTCAAAACCCAGCTGCTGCTAATTTAACCCAGCTTAATC 7080
 Qy 61 AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCys 80
 Db 7081 GCCGCCATGAATTTGCTTCAGAAATCCAAAGCCAAAGATTTTCAGAAACGGCATCGGCTGT 7140
 Qy 81 ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
 Db 7141 GTGGCTGATAAAGCACCTGAAGCGCTCAGCTTGACCGAGCTTACCAGAGCTGCCATTAAT 7200
 Qy 101 ProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMet 120
 Db 7201 CCAATGCACGCACTGAAGTTGCCAAAAAATTTGCGACACTCGCTTAAACCTTGCATG 7260
 Qy 121 LeuGluThrValAsnAlaPheIleValProThrThrArg 134
 Db 7261 CTAGAGACCGTCATGCTTTTATTGTGCCAACTACCACAGC 7302
 RESULT 3
 NMA522491 329861 bp DNA linear BCT 04-DEC-2000
 LOCUS NMA522491
 DEFINITION Neisseria meningitidis serogroup A strain 22491 complete genome;
 segment 5/7.
 ACCESSION AL162756 AL157959
 VERSION AL162756.2 GI:7380091
 KEYWORDS
 SOURCE Neisseria meningitidis 22491.
 ORGANISM Neisseria meningitidis 22491
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
 Neisseria.
 REFERENCE 1 (bases 1 to 329861)
 AUTHORS Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C., Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T., Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K., Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M., Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
 TITLE Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
 JOURNAL Nature 404 (6777), 502-506 (2000)
 MEDLINE 20222556
 PUBMED 10761919
 REFERENCE 2 (bases 1 to 329861)


```

AUTHORS      Parkhill,J.
TITLE        Direct Submission
JOURNAL      Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
              sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
              Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT      Notes:
              Details of N. meningitidis sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL, http://www.sanger.ac.uk/projects/N_meningitidis/).
FEATURES     source
              1. 329861
              /organism="Neisseria meningitidis 22491"
              /strain="22491"
              /db_xref="taxon:122587"
              /note="serogroup: A"
              194. .565
              /gene="NMA1439"
              194. .565
              /gene="NMA1439"
              /note="NMA1439, possible lipoprotein, len: 123 aa;
              unknown, contains a probable N-terminal signal sequence
              and an appropriately positioned PS00013 Prokaryotic
              membrane lipoprotein lipid attachment site"
              /codon_start=1
              /transl_table=11
              /product="putative lipoprotein"
              /protein_id="CAB84675.1"
              /db_xref="GI:7380092"
              /db_xref="SPTREMBL:Q9JU92"
              /translation="MNKTLSTLPVAILLGCAAGGNTFGSLDGGTGSGSIVKMAVE
              SOCRALNKRSWPLTALAMSAEKOAEWENKICACVQAEPNQLTGNDVMQMLDPSTR
              NQALAAITAKTVSACFKHLYR"
              212. .244
              /gene="NMA1439"
              /note="PS00013 Prokaryotic membrane lipoprotein lipid
              attachment site"
              complement(571. .580)
              /note="Core DNA uptake sequence: gccgtctgaa"
              /label=DUS
              592. .601
              /note="Core DNA uptake sequence: gccgtctgaa"
              /label=DUS
              638. .2047
              /gene="thrc"
              638. .2047
              /gene="thrc"
              /EC_number="4.2.99.2"
              /note="NMA1440, thrc, probable threonine synthase, len:
              485 aa; similar to many e.g. SW:THRC_METGL (EMBL:D14071),
              thrc, Methylobacillus glycoenes threonine synthase (EC
              4.2.99.2) (475 aa), fasta scores; E(): 0, 61.9% identity
              in 475 aa overlap. Contains Pfam match to entry PF00291
              S_T_dehydratase, Pyridoxal-phosphate dependent enzymes"
              /codon_start=1
              /transl_table=11
              /product="putative threonine synthase"
              /protein_id="CAB84676.1"
              /db_xref="GI:7380093"
              /db_xref="SPTREMBL:Q9JU91"
              /translation="MKYISTRGETAHAKPFSVLLMGLAPDGLMLPEHYPOIGREALD
              KWRLAVPELAFMLRFLVTDIPEDDLRLNRYTEAAFGCTKEITPVRTLSDGKIKQ
              ALNSGPTLAFKMAQFIGNAFEVVLNKEGKLNILGATSGDTSAAEYALRGKGVN
              VMLSPDGKMSAFQAQMYSLQDGNHNIHVKMGFDCCDQIVKAVONDAAPAEKRYHIG
              TVNSINRWIYAQVYVYPAGYFKATQSNDEQVSCVPSGNGFWCAGHIAKOMGLPVR
              RLIVATNENDVDFEFTFGAYRPNRSHTYVTSSPSISKASNFREVFVFLMDRDPQ
              EINTLWAAVAGKFNLRFDLQKVGKYGFTSGKSTHADRLATIROVYEQDELIDPH
              TADGVKVAVEREETVVCLETALAANKFDTATIREAVGDVAIPRAALEGLENLPRV
              QTVFNSADAVKGIETILA"
              875. .1780
              /gene="thrc"
              /note="Pfam match to entry PF00291 S_T_dehydratase,
              Pyridoxal-phosphate dependent enzymes, score 231.80,
              E-value 1e-65"

misc_feature
              complement(922. .931)
              /note="Core DNA uptake sequence: gccgtctgaa"
              /label=DUS
              2094. .2888
              /gene="NMA1441"
              2094. .2888
              /gene="NMA1441"
              /note="NMA1441, len: 264 aa; unknown"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein NMA1441"
              /protein_id="CAB84677.1"
              /db_xref="GI:7380094"
              /db_xref="SPTREMBL:Q9JU90"
              /translation="MTVPMQETHFSILLDELAAKOEAAPHLLADGTQVWIRKAGRH
              NARWYALLGMVARYLKLGVLPKVPKSLGGEPATIESRKLVELRAAGTAVPELLAVRK
              NALMFGNLEGIPLDTQIROEAEAGKADAWLAGLEAIARVHKKQOFLSOFARNMMSDG
              KNISFLDFEDDPSEVLTIACQARDWLCYIHSTALILKNGLLGEAAEKKWGVLSDDQ
              AEIQKLIATYKPIILPIRRLEHPWRGRDALRAASISLISLADHPP"
              2904. .2937
              /note="stem loop containing DNA uptake sequences: acgat
              gccgtctgaa gcc ttacagacgc atata"
              2909. .2918
              /note="Core DNA uptake sequence: gccgtctgaa"
              /label=DUS
              complement(2923. .2932)
              /note="Core DNA uptake sequence: gccgtctgaa"
              /label=DUS
              3065. .3068
              /label=DUS
              3076. .3852
              /gene="fpr"
              3076. .3852
              /gene="fpr"
              /EC_number="1.18.1.2"
              /note="NMA1442, fpr, probable ferredoxin--NADP reductase,
              len: 258 aa; similar to many e.g. SW:FENR_AZQVI
              (EMBL:L36319), fpr, Azotobacter vinelandii
              ferredoxin--NADP reductase (EC 1.18.1.2) (257 aa), fasta
              scores; E(): 0, 69.8% identity in 255 aa overlap. Similar
              to NMA1664, fasta scores; E(): 1.3e-30, 34.6% identity in
              257 aa overlap. Contains Pfam match to entry PF00175
              oxidored_fad, Oxidoreductase FAD/NAD-binding domain"
              /codon_start=1
              /transl_table=11
              /product="putative ferredoxin--NADP reductase"
              /protein_id="CAB84678.1"
              /db_xref="GI:7380095"
              /db_xref="SPTREMBL:Q9JRE3"
              /translation="MAAFNTOKVLSVHHTDAYFTTCTRDESLRFENGQFVMVGLMV
              DGKPLMRAYSVASANWEHLSEFFSKYKVDGDLTSRLOHLKVGDDVLISKKPTGTLVAG
              DLNPGKHLYLLSTGTGFIAPLSTKDPETIYQFKEILLVHGVRYKKDLAYYDRFTKEL
              PEHYLGLVKEKLIYPIVSRREEFEHGRHLTDLMSVSKLFDLIGLPIKINPQDDRAML
              CGSPAMLKDTCKVLDDFGLTVSPKTVGRGVYLIERAFVDQ"
              3394. .3768
              /gene="fpr"
              /note="Pfam match to entry PF00175 oxidored_fad,
              Oxidoreductase FAD/NAD-binding domain, score 20.00,
              E-value 8.4e-05"
              3866. .3875
              /label=DUS
              /note="Core DNA uptake sequence: gccgtctgaa"
              3964. .3983
              /label=DUS
              /note=">= 90% match to ATTCNNNNNNNNNGGAAT"
              /label=DUS
              complement(3984. .4038)
              /note="RS103"
              /label=BS103
              4039. .4058
              /note=">= 90% match to ATTCNNNNNNNNNGGAAT"
              /label=DUS
              4159. .4320
              /gene="NMA1443"
              4159. .4320
              /label=DUS

```



```
/gene="NM1045"
/complement(1675. .2469)
/note="NM1045"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41443.1"
/db_xref="GI:7226285"
/translations="MTLPMTQETRESILLDELAQKQAAIAPIHLADCTKYWKAGRH
NARWYALLGWARYIKLVKVPVPSLGGEPATATESKRLYELRSQAIAPPELLAYRK
NALMFGLEGIPDAQIQAEAGKAGSDAWGLAEIARVHKRQFSLQAEIARNMDSG
KNTSFDFDDPDEQVETIAQCQARDWLCYTHSTALILKNGGLLEAAAEKRWGVLSDOP
AETQKIAATVKTPIITRLEHPRWRGRDLRLAASISLISLADMP"
/complement(2515. .3924)
/gene="NM1046"
/complement(2515. .3924)
/note="NM1046"
/note="similar to SP:P23669 GB:X36037 PID:580784
PID:580991 percent identity: 71.31; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="AAF41444.1"
/db_xref="GI:7226286"
/translations="MKYISTRGETAHKPFSEVLIMGLAPDGLMLPBYPIQIGRETLD
KWGLAYPELAFETIMFLVTDIPEDLDRLNRTYTEAFGTEKITPVRTLSDGIKIQ
ALNSGTPKAFKMAQELNFAEYVLNKEGKKNILGATSGDGSAAEYALRGKGVN
VMSLDPKMSAFQRAQYSLQDENHNIYAKWGFDDCQDQIVRAVQNDAAFEKKYHLG
TVLNNNGRIYVAQVYFYFAGYFNATNSDETVCVPSGNGFNAGAHIAQKMLPIR
ELIVATNENVDLEFFKTGAYRPRNSAHTYVTSKMDISKASNERFEVFDLMDRDP
EINTLAEVAGGFDLRFALDKVGGYGTSGKSHADRLATIKQYEQDQELIDPH
TADCVKVARVEEGENVCLTALAAKFDATIREAVGDAIAPRAALEGLNLPQV
QTVPNSADAVKGIIEQTLA"
/complement(3997. .4368)
/gene="NM1047"
/complement(3997. .4368)
/note="NM1047"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41445.1"
/db_xref="GI:7226287"
/translations="MNKTLSTLPVAILLGGCAAGGNTFGSLDGGTGMGGSIVKMAVG
SOCAELDKSEWELTALAMSAEQAWEKNCACVAQEAPEMTGNDVQMLAPSTR
NQALALATKTVSACFHLNR"
/complement(4620. .6089)
/gene="NM1048"
/complement(4620. .6089)
/note="NM1048"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF41446.1"
/db_xref="GI:7226288"
/translations="MERKHGTYRDLHRPASFEATRDYELEHLOIMQPKWRPNLPER
DYREWEDLIPAMAGTIGKVMYGAFAAFAAPLGDPSFVLENRVYELLIAAFILL
VSGFELPCANLPGTHGLPIPIVWSAGGHPAFGISIAVIGLIMALFRGSGHAKL
TSNCGCGLLYLGFICTGOVKKLSWAGGNMPTIATVIIVTVYVALLHEHFWKL
WLAFLGLGIAVVAFGALGAPPEHTAPGLPPKSPAYWGCNSGHLGLPTAESFLV
FPFVAVLAWMSDFLGHQVFQKISYPEKTDKVLNIDDDMTSCSVQAVGSLGGAN
TFSWGTYIVPASIAKRPYIPGGAVLTAVICLIAGLWGYPMDLAIWQVLSVALVGYV
LPLEAGMETNRKGTQSAIIVVFSSALNVPVFGWALTMLDLNLGLTGCKERSAQLG
FAGRLVPAVGFILCVAMGAVGMLPGTPPELHFKSIG"
/complement(6345. .7259)
/gene="NM1049"
/complement(6345. .7259)
```

```
/gene="NM1049"
/note="similar to GB:M13687 SP:P23841 GB:X63976 PID:148255
PID:41595 percent identity: 48.44; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, putative"
/protein_id="AAF41447.1"
/db_xref="GI:7226289"
/translations="MDAVOLKSFVAHAEGNLTOAKRFLSQPAPVAQAQIKALEEYVG
TPLFRFGKGMVLRAGEILLPEASELLOVKHKLHEFAKTLAGDYSETSGLIHPID
SALYALITDNIGQTAPKTRHLHQYMGSEILSRIOHKTLLHGGFLGFLNNAQKRSVFL
QNTYALICQSOYPHLTSLPSQSCVWEMSGVSGSRKHLHQFWRNRLSPKKQI
LCDYPTITDIVAGIGVAVPGNKAKEAAKAGAVAIIESCRHSMPLNFTYAEYED
NPHVSLLECTEKWGVQAVQPPVVSQN"
/complement(7718. .8683)
/gene="NM1050"
/complement(7718. .8683)
/note="NM1050"
/note="similar to GB:L27157 PID:451858 percent identity:
62.86; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transposase, IS30 family"
/protein_id="AAF41448.1"
/db_xref="GI:7226290"
/translations="MSYTLQTQGERYHIQYLSRHCTEIAKQLNRHKSTISREIRRH
RTQGOQYSAKAQROQSIKORQOPYKLDLSOLIOHTLIRRLKSPVQVCAYLCKHH
QIUHSTIIVYLRQDSKNSGLWQHLRICKSPYKRYGSDWTKGVPNRVGIENRPA
IVDKSRIGDEADTVGKGKSAALLIVERVRYTIIICKLBSLAEDTAAVAALAK
AKHDVHTITMDNGKEFYQHTKITKALAEYFCRPHSWEGNLNGLIRQTFPK
QTDPRNLSREIRRVQDELNHRPKRTILGYETPSVFLNLFQPLIH"
/complement(8805. .10715)
/gene="NM1051"
/complement(8805. .10715)
/note="NM1051"
/note="similar to GB:L42023 PID:1007269 PID:1221430
PID:1205538 SP:Q57242 percent identity: 69.46; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAF41449.1"
/db_xref="GI:7226291"
/translations="MNILSVENSAFAGHVALLDKTSFQLDSGEKVGILGRNGAKSS
FLKTLAGLQKLDGQIIQNNLKIVYVYVQESFFQKDATVFTVAEGLGEIRDLRRYH
HVSHLENGSEALLKELNELEAKDGWKLDAVQKOTLGEGLPNEKIGNLSSG
OKKRLAQAAWQKPDVLLDEPTNHLIDDAIWLLENLLKAFEGSLVVTTHDRFLDN
IATRIVELDRGILRSYFGSFSKYSEKKAQELAVEAENRLEFDFKHAQEAHWIRGIEA
RRTNREGVRLEELRQRARRNVQGVNFKLDSGEKSGKIIAELEHASPAYGKVI
MDKFSAILQGRDKTIGLPGNIGIKTTLKLILGELQPTYGRIRIGSKQEVAYFOFRS
ALNENDTVFTLQGNQDYVEVGKKHVMYLEDLFHPARAQSPVSLSGGERNRLI
LAKLTPRANILVLDPTNDLIDTQELLEDLLRDYQGVFLVSHDRMFLDNVITQSI
```

Alignment Scores:

Pred. No.:	0.144	Length:	11489
Score:	116.00	Matches:	29
Percent Similarity:	46.28%	Conservative:	27
Best Local Similarity:	23.97%	Mismatches:	59
Query Match:	17.16%	Indels:	6
DB:	1	Gaps:	3

US-10-048-196-2 (1-134) x AE002455 (1-11489)

Qy	1	MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys	20
Db	4368	ATGAATAAACCTTGTCTATT-----TTGCCGGTGGCAATCTTACTCGCGGCTGC	4318
Qy	21	AlaSerThrProGluSer-----AsnProLysAsnSerSerAlaAsnLeuThrThrSer	38
Db	4317	GCCCGCGGAGCGGTAAACACATTCGCACGCTTAGACGGTGGCACGGCATGGCGGCAGC	4258
Qy	39	LeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrp	58

```

Db 4257 ATCGTCAAAATGCGGTGGAGCCCAATCGCGTGGGAATTGCACAAACGACGGAATGC 4198
||||| ||||| ||||| ||||| |||||
QY 59 LysileAlaAlaMetLysLeuSerGluSerLysAlaLysLysSerGluThrAlaCys 78
||||| ||||| ||||| ||||| |||||
Db 4197 CGTTTACCGCGTGGCGATGAGTCGCGAAACACGCGGAGTGGGAAACAGATTTCG 4138
||||| ||||| ||||| ||||| |||||
QY 79 GlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAla 98
||||| ||||| ||||| ||||| |||||
Db 4137 GCTTGGTGGCGCCCAAGACACCGGACGATGACCGGCAACGATGTG---ATGCAGATG 4081
||||| ||||| ||||| ||||| |||||
QY 99 IleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysPro 118
||||| ||||| ||||| ||||| |||||
Db 4080 CTGGCTCCGTCACCGCAATCAGCACTTGGCGCCCTGACCGCCCAAAACGGTTTCCGCC 4021
||||| ||||| ||||| ||||| |||||
QY 119 Cys 119
|||||
Db 4020 TGC 4018
|||||

RESULT 5
AX044031/c
LOCUS
DEFINITION
Sequence 110 from Patent WO0066791.
ACCESSION
AX044031
VERSION
AX044031.1 GI:11342915
KEYWORDS
Neisseria meningitidis.
ORGANISM
Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 349980)
AUTHORS
Pizza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J. C.,
Masignani, V., Galeotti, C., Mora, M., Ratti, G., Scariselli, M.,
Scarlato, V., Rappuoli, R., Frazer, C. M. and Grandi, G.
TITLE
Neisseria genomic sequences and methods of their use
JOURNAL
Patent: WO 0066791-A 110 09-NOV-2000;
FEATURES
LOCATION/Qualifiers
source
1. 349980
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
/note="sequence too long, cut in 8 pieces, -seq 1: 1 to
349980 349980 bases-seq 108: 300001 to 349980 349980
bases-seq 109: 600001 to 349980 349980 bases-seq 110:
900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
2272325 172325 bases"
BASE COUNT 81351 a 86755 c 95584 g 86290 t
ORIGIN

Alignment Scores:
Pred. No.: Length: 10.2 349980
Score: 116.00 Matches: 29
Percent Similarity: 46.28% Conservative: 27
Best Local Similarity: 23.97% Mismatches: 59
Query Match: 17.16% Indels: 6
DB: 6 Gaps: 3

US-10-048-196-2 (1-134) x AX044031 (1-349980)

QY 1 MetMetLysIleLeuTyValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
||||| ||||| ||||| ||||| |||||
Db 164197 ATGAATAAACCTTGCTATT-----TTCCCGGTGGCAATCTTACTCGGCGCTGC 164147
||||| ||||| ||||| ||||| |||||
QY 21 AlaSerThrProGluSer-----AsnProLysAsnSerSerAlaAsnLeuThrThrSer 38
||||| ||||| ||||| ||||| |||||
Db 164146 GCGCGCGGAGCGGTAAACACATTCGCACCTTAGACGGTGGCACAGCGATGGCGGCACG 164087
||||| ||||| ||||| ||||| |||||
QY 39 LeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyTrp 58
||||| ||||| ||||| ||||| |||||
Db 164086 ATCGTCAAAATGCGGTGGGACCAATCGCGTGGGAATTGCACAAACGACGGAATGC 164027
||||| ||||| ||||| ||||| |||||

```

```

QY 59 LysileAlaAlaMetLysLeuSerGluSerLysAlaLysLysSerGluThrAlaCys 78
||||| ||||| ||||| ||||| |||||
Db 164026 CGTTTACCGCGTGGCGATGAGTCGCGAAACACGCGGAGTGGGAAACAGATTTCG 163967
||||| ||||| ||||| ||||| |||||
QY 79 GlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAla 98
||||| ||||| ||||| ||||| |||||
Db 163966 GCTTGGTGGCGCCCAAGACACCGGACGATGACCGGCAACGATGTG---ATGCAGATG 163910
||||| ||||| ||||| ||||| |||||
QY 99 IleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysPro 118
||||| ||||| ||||| ||||| |||||
Db 163909 CTGGCTCCGTCACCGCAATCAGCACTTGGCGCCCTGACCGCCCAAAACGGTTTCCGCC 163850
||||| ||||| ||||| ||||| |||||
QY 119 Cys 119
|||||
Db 163849 TGC 163847
|||||

RESULT 6
AP000686
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone CMB9-99M9 map 11q25, WORKING DRAFT
SEQUENCE, 12 unordered pieces.
ACCESSION
AP000686
VERSION
AP000686.2 GI:8118874
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens DNA, clone:CMB9-99M9.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 68466)
AUTHORS
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE
Homo sapiens 68,466 genomic DNA of 11q25
JOURNAL
Published Only in DataBase (1999)
REFERENCE
2 (bases 1 to 68466)
AUTHORS
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE
Direct Submission
JOURNAL
Submitted (08-NOV-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan [E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924]
COMMENT
On May 31, 2000 this sequence version replaced gi:6997560.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: CMB9-99M9
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 64876 bases at least Q40
Consensus quality: 66248 bases at least Q30
Consensus quality: 66996 bases at least Q20
Insert size: 67366; sum-of-contigs
Quality coverage: 8.74x in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently consists of
12 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 14889 contig of 14889 bp in length
14990 27066 contig of 12077 bp in length
27167 34495 contig of 7329 bp in length
34596 41662 contig of 7067 bp in length

```



```

/note="assembly_name:Contig19"
61440. .77318
misc_feature
/note="assembly_name:Contig20"
77419. .98834
misc_feature
/note="assembly_name:Contig21"
98935. .122184
misc_feature
/note="assembly_name:Contig22"
12285. .152508
misc_feature
/note="assembly_name:Contig23
clone_end:T7
vector_side:right"
152609. .187087
misc_feature
/note="assembly_name:Contig24"
53713 a 41318 c 39330 g 51421 t 1305 others
BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 94.9 Length: 187087
Score: 103.50 Matches: 39
Percent Similarity: 36.88% Conservative: 13
Best Local Similarity: 27.66% Mismatches: 52
Query Match: 15.31% Indels: 37
DB: 2 Gaps: 5

```

```

best Local Similarity: 27.60% mismatches: 32
Query Match: 15.31% Indels: 37
DB: . 2 Gaps: 5

US-10-048-196-2 (1-134) x AC079123 (1-187087)

Qy      2 MetLysIleLeuYrValThrAlaThrLeuMetThrAlaPheThrLeuAla----- 18
          TTAAGGCTGATGCATCACGCCACAGTAAATGCAACTGTCTTCAGCAGCATAGAAGTG 182944
          ::::::::::::::::::::: 3:: 111

Db 182885 TTAAGGCTGATGCATCACGCCACAGTAAATGCAACTGTCTTCAGCAGCATAGAAGTG 182944
          ::::::::::::::::::::: 3:: 111

Qy      19 -----SerCysAlaSerThrProGluSer 26
          ::::::::::::::::::::: 111111

Db 182945 ACACATGGCATGGGTCAGAGTGGAATATGCAGGAAGTCTCGCTTGACATTCTCATCTG 183004
          ::::::::::::::::::::: 111111

Qy      27 AsnProLysAsnSerSerAlaAsnLeuThrSertLeuile-----LysHisAlaVal 44
          ::::::::::::::::::::: 111111

Db 183005 TATCCTTTACCAG-----ACGACCATATCTGTTCCGGGGGAACAATGCGCAGC 183049
          ::::::::::::::::::::: 111111

Qy      45 LysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTrpLysIleAlaAlaMetLys 64
          ::::::::::::::::::::: 111111

Db 183050 AAAGAGACAGCTGCTACCTTTATGGAAGT-----CCC 183082
          ::::::::::::::::::::: 111111

Qy      65 LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAaspLys 84
          ::::::::::::::::::::: 111111

Db 183083 TTACAGAGCCCTCTAAAGCCAAGTGGGTCCCCAACCTCTGCTCTATCAGAGGAGACAGT 183142
          ::::::::::::::::::::: 111111

Qy      85 AlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleasnProAspAlaArg 104
          ::::::::::::::::::::: 111111

Db 183143 GCACCTCTCAGCTGFA-----GTGAACACAGATCTCTGCCAACCATAGA 183184
          ::::::::::::::::::::: 111111

```

Db	183143	GCACCCCTCAGCTGTA-----GTGAACACAGATCTCTGCCAACCATAGA	183184
Qy	105	ThrGluValAlaGlnLysIleValArgHisserLeuLysProCysMetLeuGluThrVal	124
Db	183185	GGGAATTCCTTTTACACTAGTAGTACCCACCAAGAAGGAATGTTTCATTTTACACCC	183244
Qy	125	Asn 125	
Db	183245	AAT 183247	
RESULT	8		
LOCUS	AP003327		
DEFINITION	Homo sapiens genomic DNA, chromosome 11q, clone:Rp11-507F16, complete sequence.		
ACCESSION	AP003327		
VERSION	AP003327.2	GI:19263043	
KEYWORDS	HTG.		
SOURCE	Homo sapiens DNA, clone:Rp11-507F16.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiwara,A., Yada,T., Toki,Y., Watanabe,H. and Sakaki,Y.		

[illegible]

primer_bind	/note="N-terminal primer region for PCR"
	763..780
BASE COUNT	260 a 246 c 119 g 155 t
ORIGIN	/note="C-terminal primer region for PCR"

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
-----
Project Information
-----
Center project name: M_BA0042113
-----
Summary Statistics
-----
Sequencing vector: MI3; 0x
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 240797 bases at least Q40
Consensus quality: 242108 bases at least Q30
Consensus quality: 242698 bases at least Q20
Insert size: 204000; agarose-fp
Insert size: 248995; sum-of-contigs
Quality coverage: 10.08 in Q20 bases; agarose-fp
Quality coverage: 8.91 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0042113

misc_feature 1186..3097 /note="assembly_name:Contig33"
misc_feature 3198..5110 /note="assembly_name:Contig84"
misc_feature 5211..7669 /note="assembly_name:Contig85"
misc_feature 7770..13621 /note="assembly_name:Contig86"
misc_feature 13722..19656 /note="assembly_name:Contig87"
misc_feature 19757..31061 /note="assembly_name:Contig88"
misc_feature 31162..40106 /note="assembly_name:Contig89"
misc_feature 40207..50613 /note="assembly_name:Contig90"
misc_feature 50714..61952 /note="assembly_name:Contig91"
misc_feature 62053..74978 /note="assembly_name:Contig92"
misc_feature 75079..92714 /note="assembly_name:Contig93"
misc_feature 92815..124084 /note="assembly_name:Contig94"
misc_feature 124185..151377 /note="assembly_name:Contig95"
misc_feature 151478..178112 /note="assembly_name:Contig96"
misc_feature 178213..208382 /note="assembly_name:Contig97"
misc_feature 208483..249448 /note="assembly_name:Contig98"
misc_feature 249549..249622 /note="assembly_name:Contig99"
misc_feature 249723..249846 /note="assembly_name:Contig24"
misc_feature 249723..249846 /note="assembly_name:Contig29"
BASE COUNT 67137 a 56658 c 55913 g 68327 t 1811 others
ORIGIN

Alignment Scores:
Pred. No.: 1.34e+03 Length: 249846
Score: 94.00 Matches: 43
Percent Similarity: 42.68% Conservative: 24
Best Local Similarity: 27.39% Mismatches: 42
Query Match: 13.91% Indels: 48
DB: 2 Gaps: 10

US-10-048-196-2 (1-134) x AC125227 (1-249846)

QY 3 LysileLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSer 22
Db 25802 CGCATACAATGTGTGACTGCCACACTGCTGCGCTGGG-----GCCAGCTCCTCTCG 25855
QY 23 ThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLysHis 42
Db 25856 GTCATCTCTACCCAGGCTAGGAGCGCACAGCG-----AAACAC 25894
QY 43 AlavalysGlnThrGlnThrGlnLeuThr----- 53
Db 25895 -----TCCACAGACACAGAAACAATCCGGGTAGAGATCTGAGATGGAAC 25939
QY 54 -----GlyHisGln-----TyrTrpLysIleAlaAlaMetLys 64
Db 25940 GGAGGTACCTTGGGCGCACAGGAGGAGCTCCTGACCTGGGCACATAGAAAGTGTGAGA 25999
QY 65 LeuSerSerGluSerLysAlaLys-----IleSerGluThrAlaCys 78
Db 26000 ---CAGTCTGAAGCTTAGTAAAGTCCCAACAGACATGGTCAAGATGCGCTGT 26056
QY 79 GlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAla 98
Db 26057 GGATCTAAGCCTTGAGAGTCTCAGAGGCGCTTGGGCCCAACTGAGTTAAGTACAGAGCT 26116

QY 99 IleAsnPro-----AsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeu 116
Db 26117 GTCAGAGACAGTTTCAGCTCATACACACATCTTGGCTGAACTGGGAAGA----- 26167
QY 117 LysProCysMetLeuGluThrValAsnAlaPheIleValProThrThrThr 133
Db 26168 ACTCCTTGTGTTG-----GGAGCTTGTGTGAGCCACAGACATCC 26206
RESULT 13
AF223572 AF223572 Homo sapiens 632 bp DNA linear PRI 16-APR-2000
LOCUS Homo sapiens microsatellite D10 sequence.
DEFINITION AF223572
ACCESSION AF223572
VERSION AF223572.1 GI:7576885
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 632)
AUTHORS Pujana.M.A., Volpini.V. and Estivill.X.
TITLE Direct cloning of polymorphic CAG repeats
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 632)
AUTHORS Pujana.M.A., Volpini.V. and Estivill.X.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-2000) Molecular Genetics, Medical and Molecular
Genetics Centre-IRO, Hospital Duran i Reynals, Autovia de
Castelldefels km 2.7, L'Hospitalet de Llobregat, Barcelona 08907,
Spain
FEATURES
source 1..632
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
repeat_region 429..515
/note="microsatellite D10"
/rpt_type=tandem
/rpt_unit=cag
BASE COUNT 166 a 215 c 133 g 114 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 1.12 Length: 632
Score: 92.50 Matches: 29
Percent Similarity: 54.67% Conservative: 12
Best Local Similarity: 38.67% Mismatches: 29
Query Match: 13.68% Indels: 5
DB: 9 Gaps: 2
US-10-048-196-2 (1-134) x AF223572 (1-632)
QY 9 AlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer---Asn 27
Db 138 TCTACTGTCCTCCAGCGCCACTAGTTTACCTTCTGTTGCCAGCAGCTCCCGCAGCTCCAAAC 197
QY 28 ProLysAsnSerSerAlaAsn-LeuThrThrSerLeuIleLysHisAlaValLysGlnTh 47
Db 198 CTGCGAAGCTCACCAGCAAACTGTGCTGCCAGTCCCTCAAACTCCAAACAGCCAC 257
QY 47 rCysGlnThrGlnLeuThrGlyHis-GlnTyrTrpLysIleAlaAlaMetLysLeuSerS 67
Db 258 ACTCCAGGCCCAAGCTC-----CACCTCGGCGCTGGAATGTTATCTCCTGAATCCGCA 311
QY 67 erGluSerLysAlaLysIleSerGluThrAlaCysGlyCys 80
Db 312 CGAGTGACAGTGGCCGGTTCAGCGTCAGGCGCTGTGGCTGT 352
RESULT 14
AC097464/c AC097464 174410 bp DNA linear PRI 01-MAR-2002
LOCUS Homo sapiens BAC clone RP11-21M10 from 4, complete sequence.
DEFINITION AC097464
ACCESSION AC097464

VERSION AC097464.3 GI:16418223

KEYWORDS HTG.

SOURCE Homo sapiens.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 174410)

Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

PUBMED

9847074

2 (bases 1 to 174410)

Tomlinson, C., Haglund, K. and Spalding, L.

The sequence of Homo sapiens BAC clone RP11-21M10

Unpublished (2001)

3 (bases 1 to 174410)

Waterston, R.H.

Direct Submission

Submitted (18-OCT-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 174410)

Waterston, R.H.

Direct Submission

Submitted (25-OCT-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 174410)

Waterston, R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Oct 25, 2001 this sequence version replaced gi:16303526.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@watson.wustl.edu

----- Summary Statistics

Center project name: H_NH0021M10

Drafting Center: WIDR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frenken, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-34211; the clone sequenced to the right is RP11-328N8. Actual start of this clone is at base position 1 of RP11-21M10; actual end is at base position 174410 of RP11-21M10.

Data from AC044906 and AC040995 was used to finish this clone. AC019343. Polymorphisms have been identified between AC044906, AC040995 and AC019343.

The sequence of AC019343 has been incorporated into AC097464.

FEATURES

Location/Qualifiers

1..174410

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-21M10"

/clone_lib="RPCI-11"

321..624

/rpt_family="Alu"

997..1566

/note="similar to EST BG571049 (NID:g13578702)"

1192..1244

/rpt_family="L2"

1330..1403

/rpt_family="MIR"

1834..1865

/rpt_family="(T)n"

1836..2147

/rpt_family="Alu"

4337..4853

/note="similar to EST AW173298 (NID:g6439246) xj85g01.xl"

5234..5886

/note="similar to EST BG174121 (NID:g12680824)"

5947..5968

/rpt_family="(TGAA)n"

6046..6135

/rpt_family="MIR"

6733..6825

/rpt_family="LI"

7097..7198

/rpt_family="(GAAA)n"

7202..7331

/rpt_family="LI"

7379..7463

/rpt_family="Mariner"

8019..8067

/rpt_family="AT-rich"

8695..9214

/note="similar to EST AW117585 (NID:g6086169) xd93f02.xl"

9161..9187

/rpt_family="AT-rich"

10323..10487

/rpt_family="MER53"

10737..10809

/rpt_family="(TC)n"

11811..11936

/rpt_family="MIR"

12228..12367

/rpt_family="Alu"

12826..13022

/rpt_family="MIR"

13099..13266

/rpt_family="MIR"

13373..13664

/rpt_family="Alu"

13695..13766

/rpt_family="CRI"

14001..14312

/rpt_family="Alu"

14575..15009

/rpt_family="LI"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```
repeat_region 15376..16062
/rpt_family="L1"
repeat_region 15812..15832
/rpt_family="AT-rich"
repeat_region 16216..16314
/rpt_family="MIR"
repeat_region 16389..16411
/rpt_family="(TG)n"
repeat_region 17304..17357
/rpt_family="(TCA)n"
repeat_region 18029..18060
/rpt_family="(TG)n"
repeat_region 18964..19002
/rpt_family="(CA)n"
repeat_region 19289..19312
/rpt_family="(T)n"
repeat_region 19715..19863
/rpt_family="MIR"
repeat_region 19872..19930
/rpt_family="MaLR"
repeat_region 19945..20334
/rpt_family="MaLR"
repeat_region 20382..20473
/rpt_family="MERL_type"
repeat_region 20789..21084
/rpt_family="L1"
repeat_region 21085..21362
/rpt_family="Alu"
repeat_region 21363..23408
/rpt_family="L1"
repeat_region 22229..22283
/rpt_family="A-rich"
repeat_region 23117..23159
/rpt_family="(CA)n"
repeat_region 24431..24675
/rpt_family="MIR"
repeat_region 26324..26348
/rpt_family="(TGGGA)n"
repeat_region 26605..27008
/rpt_family="L2"
repeat_region 27080..27269
/rpt_family="MERL_type"
repeat_region 27806..27894
/rpt_family="MIR"
misc_feature 27949..28439
/Note="similar to EST BF356711 (NID:g11315785)"
repeat_region 29739..30069
/rpt_family="L1"
repeat_region 30318..30660
/rpt_family="MERL_type"
misc_feature 30675..31052

Alignment Scores:
Pred. No.: 1-23e+03 Length: 174410
Score: 92.50 Matches: 29
Percent Similarity: 54.67% Conservative: 12
Best Local Similarity: 38.67% Mismatches: 29
Query Match: 13.68% Indels: 5
DB: 9 Gaps: 2

US-10-048-196-2 (1-134) x AC097464 (1-174410)

QY 9 AlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer---Asn 27
Db 99383 TCTACTGTCTCCAGCGCCACTAGTTTACCTTCGTGGCCAGCAGCTCCCGAGCTCCAAAC 99324

QY 28 ProLysAsnSerSerAlaAsn-LeuThrThrSerLeuLeuLeuHisAlaValLysGlnTh 47
Db 99323 CCGTGAAGCTCCAGCAAGAACTGTGTGTCAGTCCCTCAAACTCCAAACCAAGCCAC 99264

QY 47 rCysGlnThrGlnLeuThrGlyHis-GlnTyrrTrpLysLeuAlaAlaMetLysLeuSerS 67
Db 99263 ACTCCAGGCCAAGCTC-----CACCTCGCGCTGGAAATGGTTATCTCTCTGATCCGGCA 99210
```

```
QY 67 erGluSerLysAlaLysIleSerGluThrAlaCysGlyCys 80
Db 99209 GCAGTGACAGTGGCGGTTTCAGGCTCAGGGCGCTGGCGTGT 99169

RESULT 15
AL845171/c
LOCUS
DEFINITION
Mus musculus chromosome 4 clone RP23-38N8, *** SEQUENCING IN
PROGRESS ***
ACCESSION
AL845171
VERSION
AL845171.2 GI:22204867
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
house mouse.
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 226325)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Sims,S.
Direct Submission
Submitted (31-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 11, 2002 this sequence version replaced gi:22033729.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
----- Project Information
Center project name: BM38N8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 218645 bases at least Q40
Consensus quality: 221861 bases at least Q30
Consensus quality: 223723 bases at least Q20
Insert size: 225125; sum-of-contigs
Insert size: 225774; 1.7% error; agarose-fp
Quality coverage: 5.27x in Q20 bases; sum-of-contigs Quality
coverage: 5.29x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 10850: contig of 10850 bp in length
* 10851 10950: gap of 100 bp
* 10951 40598: contig of 29648 bp in length
* 40599 40698: gap of 100 bp
* 40699 88025: contig of 47327 bp in length
* 88026 88125: gap of 100 bp
* 88126 103225: contig of 15100 bp in length
* 103226 103325: gap of 100 bp
* 103326 117128: contig of 13803 bp in length
* 117129 117228: gap of 100 bp
* 117229 1220412: contig of 3184 bp in length ;
* 120413 120512: gap of 100 bp
* 120513 145346: contig of 24834 bp in length
* 145347 145446: gap of 100 bp
* 145447 164482: contig of 19036 bp in length
* 164483 164582: gap of 100 bp
* 164583 172295: contig of 7713 bp in length
* 172296 172395: gap of 100 bp
* 172396 179689: contig of 7294 bp in length
* 179690 179789: gap of 100 bp
* 179790 200307: contig of 20518 bp in length
* 200308 200407: gap of 100 bp
* 200408 220883: contig of 20476 bp in length
```

Db 114239 ATATTCTCTAGAGTCAGTGTAACCTGCTATTTTGAGCAATGAATCTTCTAGCCTTACAATT 114180

sections only once, except for a 100 base overlap.
The true left end of clone RP11-550P23 is at 61944 in this
sequence. The true right end of clone RP11-374F3 is at 100 in this
sequence.

FEATURES

source	Location/Qualifiers	
repeat_region	1..62043	13226..13361
repeat_region	/organism="Homo sapiens"	/note="MER69 repeat: matches 507..654 of consensus"
repeat_region	/db_xref="taxon:9606"	13361..13550
repeat_region	/chromosome="13"	/note="MER69 repeat: matches 2224..2511 of consensus"
repeat_region	/clone="RP11-223E19"	13576..13841
repeat_region	/clone_lib="RPC1-11.1"	/note="AluX repeat: matches 19..299 of consensus"
repeat_region	38..454	16105..16701
repeat_region	/note="MER66A repeat: matches 97..486 of consensus"	/note="L1MB1 repeat: matches 5555..6156 of consensus"
repeat_region	504..837	16702..16997
repeat_region	/note="L1ME1 repeat: matches 5763..6112 of consensus"	17245..17551
repeat_region	838..871	/note="AluX repeat: matches 1..297 of consensus"
repeat_region	/note="MLT1B repeat: matches 326..359 of consensus"	18666..18789
repeat_region	872..1149	/note="L2 repeat: matches 2623..2750 of consensus"
repeat_region	/note="AluX repeat: matches 1..285 of consensus"	18873..18989
repeat_region	1150..1481	/note="FLAM_C repeat: matches 1..117 of consensus"
repeat_region	/note="MLT1B repeat: matches 1..326 of consensus"	19209..19507
repeat_region	1672..1974	/note="AluSq repeat: matches 1..296 of consensus"
repeat_region	/note="AluX repeat: matches 5..306 of consensus"	19592..19629
repeat_region	2834..3059	/note="19 copies 2 mer aa 78% conserved"
repeat_region	/note="MIR repeat: matches 20..250 of consensus"	20154..20422
repeat_region	3192..3482	/note="AluX repeat: matches 5..292 of consensus"
repeat_region	3551..3709	/note="AluX repeat: matches 1..133 of consensus"
repeat_region	/note="L1MD3 repeat: matches 7566..7729 of consensus"	21501..21634
repeat_region	3712..3862	/note="L2 repeat: matches 2705..2750 of consensus"
repeat_region	/note="AluSq/x repeat: matches 1..151 of consensus"	22047..22090
repeat_region	3863..3902	/note="MIR repeat: matches 16..252 of consensus"
repeat_region	/note="20 copies 2 mer aa 77% conserved"	23518..23726
repeat_region	3905..4207	/note="AluX repeat: matches 1..310 of consensus"
repeat_region	/note="AluX repeat: matches 2..303 of consensus"	25476..25718
repeat_region	4293..4364	/note="MER20 repeat: matches 1..218 of consensus"
repeat_region	/note="MIR repeat: matches 47..120 of consensus"	25898..25932
repeat_region	/note="HAL1 repeat: matches 1..588 of consensus"	/note="7 copies 5 mer aaag 91% conserved"
repeat_region	5830..6144	26425..26844
repeat_region	/note="AluX repeat: matches 1..303 of consensus"	/note="L1MD3 repeat: matches 7298..7739 of consensus"
repeat_region	6145..6196	26933..27243
repeat_region	/note="26 copies 2 mer aa 71% conserved"	/note="AluX repeat: matches 1..311 of consensus"
repeat_region	6284..6414	27244..27305
repeat_region	/note="AluJb/FLAM repeat: matches 24..122 of consensus"	/note="31 copies 2 mer at 69% conserved"
repeat_region	6418..6735	27314..27413
repeat_region	/note="AluYb8 repeat: matches 1..315 of consensus"	/note="AluY repeat: matches 2..69 of consensus"
repeat_region	6931..7221	27470..28009
repeat_region	/note="AluSp repeat: matches 1..292 of consensus"	/note="Cpg Island"
repeat_region	7336..7644	/evidence="not experimental"
repeat_region	/note="AluSq repeat: matches 1..310 of consensus"	/note="Sequence from uni-directional dGTP big dye terminator reads only."
repeat_region	7645..7677	28333..28424
repeat_region	/note="11 copies 3 mer gga 90% conserved"	/note="L2 repeat: matches 2604..2705 of consensus"
repeat_region	7694..8059	28593..28769
repeat_region	/note="L1MD3 repeat: matches 7359..7739 of consensus"	/note="L2 repeat: matches 2546..2709 of consensus"
repeat_region	9512..9798	29645..29931
repeat_region	/note="AluX repeat: matches 3..296 of consensus"	/note="AluX repeat: matches 1..289 of consensus"
repeat_region	9924..10230	29937..29984
repeat_region	/note="AluJb repeat: matches 1..306 of consensus"	/note="3 copies 16 mer 100% conserved"
repeat_region	10380..10692	30857..31159
repeat_region	/note="AluX repeat: matches 1..311 of consensus"	/note="AluX repeat: matches 1..304 of consensus"
repeat_region	10694..10852	31171..31491
repeat_region	/note="MIR repeat: matches 8..186 of consensus"	/note="AluJb repeat: matches 1..312 of consensus"
repeat_region	11112..11315	31524..31924
repeat_region	/note="MER20 repeat: matches 1..218 of consensus"	/note="L2 repeat: matches 2251..2674 of consensus"
repeat_region	12006..12206	32198..32505
repeat_region	/note="MIR repeat: matches 35..237 of consensus"	/note="AluX repeat: matches 9..306 of consensus"
repeat_region	12474..12785	32670..32783
repeat_region	/note="MER69 repeat: matches 2..335 of consensus"	/note="FLAM_A repeat: matches 1..126 of consensus"
repeat_region	12786..13082	33530..33839
repeat_region	/note="AluJb repeat: matches 3..298 of consensus"	/note="AluY repeat: matches 1..309 of consensus"
repeat_region	13083..13146	34011..34386
repeat_region	/note="MER69 repeat: matches 335..395 of consensus"	/note="THE1C repeat: matches 1..371 of consensus"
		34427..35448
		/note="L1PA16 repeat: matches 5155..6157 of consensus"
		35475..35922

Db 762 GGCGTTTCCCAACCATCTCAGCA 785

RESULT 18
LMFL163/c

LOCUS
LMFL163
Leishmania major Friedlin chromosome 21 cosmid L163.
AL359778
VERSION
AL359778.1 GI:8744978
KEYWORDS
chromosome assembly protein; synaptobrevin-like protein.
SOURCE
Leishmania major.
ORGANISM
Leishmania major.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE
1 (bases 1 to 27081)
Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and Smith, D.F.
AUTHORS
A physical map of the Leishmania major Friedlin genome
Genome Res. 8 (2), 135-145 (1998)
9477341
PUBMED

TITLE
2 (bases 1 to 27081)
Hilbert, H., Wedler, H., Wedler, E., Duesterhoeft, A., Ivens, A.C., Quail, M., Rajandream, M.A. and Barrell, B.G.
AUTHORS
Direct Submission
Submitted (21-JUN-2000) European Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, (E-mail: barrell@sanger.ac.uk) and GATC GmbH, Fritz-Arnold-Str 23, D-78467 Konstanz, Germany
see <http://www.ebi.ac.uk/parasites/leish.html>
Notes:
Details of leishmania sequencing at the Sanger Centre are available on the World Wide Web.
see <http://www.sanger.ac.uk/Projects/Lmajor/>
CDS are numbered using the following system eg L163.01. L163 (cosmid name). 01 (first CDS)
To make the cosmid library Leishmania major Friedlin DNA was partially digested with Sau3AI prior to cloning into BamHI site of the cosmid shuttle vector cGHV (Ryan et al. 1993) Gene 131:145-150). The sequence of the packaged vector was determined by Peter Wyler and Ken Stuart at Seattle Biomedical Research Institute, and is available as accession number U59231.
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.

Gene prediction is done using:
(1) the FramePlot program of Bibb et al.,
Gene 30:157-166(1984) as implemented
at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. (2)
codon preference based on the codon usage table for Leishmania at <http://www.kazusa.or.jp/codon/>,
(3)
the Hexamer program which was written by Richard Durbin as an integral part of the ACEDB-based analysis tools for the C.elegans Genome Sequencing Project. The program calculates the log-likelihood score for a given DNA segment based on the frequency of 6-mers, normalised for the base-pair composition of the genome. The program was trained using a fasta file of confirmed Leishmania major coding sequences (CDS), i.e. from ATG start codon to the stop codon.
CAUTION: We may not have predicted the correct initiation codon. where possible we choose an initiation codon (atg) which is preceded by a stretch of pyrimidines or part of a Kozak sequence. If this cannot be identified we choose the most upstream initiation codon.
Transmembrane domains were predicted as implemented at the TMHMM server: <http://www.cbs.dtu.dk/services/TMHMM-1.0/>
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence

overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid L163 is overlapped at the 5' end by L6294 (AL354533).
THIS ANNOTATION IS PRELIMINARY.

FEATURES
Location/Qualifiers
1..27081
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="21"
/clone="cosmid L163"
/complement(1..2505)
/gene="L163.04"
/complement(<1..2505)
/note="L163.04"
/note="L163.04, len > 834 aa, unknown"
/label="L163.04"
/codon_start=1
/product="hypothetical protein L163.04"
/protein_id="CAB95267.1"
/db_xref="GI:8744979"
/translation="MAYGSPANPHLAALAKDVHRTITGNISGPERKGEYTMVLOOFD
MLMATSVEEVAAMAVECREEMRELDLQSRNVLCPDLFRLHRLTTPGRCALGVNGY
SLPSTVSDDTYQILLEQIEDAWQVLLASMRQVPEDEQHEEIVNDFERLEERC
SDYLRAMLODTMSAVHLLDITVMDASELSTQVYCAALMKHPPDPVLAALVIEYKQ
RSIPUSLAAAEQLERYLRFMKGEAADVFESVTAAGVPLSRVFAFLNLTIDANA
VLDAYDLMDYGIAPDESTLRILCKQSRMEYAKLHFTTVLKNMATATSAQNAHSPSP
PPRPLASGGAGVAPVHTSAAATSTASTASRRDRSRYFAQRIEELAEKPKASILE
ALKVLHRAEVECTSLGGDTYINSALLRHVCRSTTPRHVLLVLPFSSLAYRPGVSDGV
RAASVEGGSKSGHPAGDFDPPANEELLEGEKSAVNEELGELGPGATSRGG
EPEAKGRPRPRQDPYFADSSSEGTYLEVLSYSPNSVQLQALSAQYEGQVPH
MEVASAADAFECMGASHGTMSPVDCRFLHALGYIYIDNRWRDRAHLVRLVLEMFY
AYHVFRLADLSGTEKFSYFAVIGTKTGLDGLGFLFAASLALDKTVNPKVATKLT
AKRDSHVIETAELEKCGWDYHRLFCDLMLPEQASAEKESPAAGANGACSRD
HHHGRGANGSHAGAVASAAAAGEHLEKESIYTSAGQGFSSYNNSHLLSARPSLVT
YMYALDGLVDKKTADWLDRKLAARQALASKRTR"
/complement(6478..7809)
/gene="L163.05"
/complement(6478..7809)
/note="L163.05"
/note="L163.05, len = 442 aa, unknown; uncertain
translation start point, may use M120"
/codon_start=1
/label="L163.05"
/product="hypothetical protein L163.05"
/protein_id="CAB95268.1"
/db_xref="GI:8744980"
/translation="MDVPSRPSVAASSTAPOGRPKPRALALORSGGALGSGRTORGE
SRVSGTAIHGAPTRQPAHSERSGORHTSSGTSLSIAANNAAECSAPALPSTGCPAS
SAQOHGASVDLPYALHMDHEHFTNRNNKTEELPGLVNSLLEGTLEAPYTHLWA
QAYASPADLLPAKAKASVLPDMFVRESQQTNGFYACVRCGTCPVSPHQVIFASAC
SLRGIAVFADLHMNGVELRVCTPTVKGAPQSPRQSSSPFPAVKSLGKVGGAEL
DVAGGLRFLVHCNCLGVNREGEITAPAAAPATTLFCANACSLVYPVYTRARL
DQSIWTDGADPEAGSGRGGASMFGRASAEHAQSRATLGLSLFGPHGAVANGVE
GAIGVDYREEGIIISYAGDDRLDASFDALLKDLDPCADLPISSSVSGD"
/complement(10203..10976)
/gene="L163.06"
/complement(10203..10976)
/note="L163.06, len = 256 aa, possibly synaptobrevin-like
protein.; carboxy terminus exhibits reasonable similarity
to several, e.g. CAB1004 synaptobrevin-like protein (240
aa, Arabidopsis thaliana, EMBL: AL132957, CAB71004); Fasta
scores: E():3.3e-09, 43.2% identity in 81 aa"
/codon_start=1
/label="L163.06"
/product="possible synaptobrevin-like protein"
/protein_id="CAB95269.1"
/db_xref="GI:8744981"
/translation="MASVNEESHALYGAIVVRLVDRVLMCKTPSAPMDGTIPSTAW
ADLVSRCSAPHRTSAFNTADKPKQEVTLVSHLMTDDAFGYGIGAKAVSRDGH
AALDELAALFKKMFVEPSTNPKLADVFPARDLMLKYSGAAGSANKVKKVL
AVDEKKNALDNVERVIORGQRIDIVQATDDLQFQAEGRSSRDLRNQMWSSMKG

Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 184024)
Worley, K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 184024)
Worley, K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:21744043.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GURJ
Center clone name: CH230-173B13
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 139590 bases at least Q40
Consensus quality: 145742 bases at least Q30
Consensus quality: 150380 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1023: contig of 1023 bp in length.
* 1024 1123: gap of unknown length
* 1124 2650: contig of 1527 bp in length
* 2651 2750: gap of unknown length
* 2751 3860: contig of 1110 bp in length
* 3861 3960: gap of unknown length
* 3961 5238: contig of 1278 bp in length
* 5239 5338: gap of unknown length
* 5339 6445: contig of 1107 bp in length
* 6446 6545: gap of unknown length
* 6546 7819: contig of 1274 bp in length
* 7820 7919: gap of unknown length
* 7920 9222: contig of 1303 bp in length
* 9223 9322: gap of unknown length
* 9323 10916: contig of 1594 bp in length
* 10917 11016: gap of unknown length
* 11017 12286: contig of 1270 bp in length
* 12287 12386: gap of unknown length
* 12387 13954: contig of 1568 bp in length
* 13955 14054: gap of unknown length
* 14055 15098: contig of 1044 bp in length
* 15099 15198: gap of unknown length
* 15199 17232: contig of 2034 bp in length
* 17233 17332: gap of unknown length
* 17333 19130: contig of 1798 bp in length
* 19131 19230: gap of unknown length
* 19231 20390: contig of 1160 bp in length

20391 20490: gap of unknown length
20491 22099: contig of 1609 bp in length
22100 22199: gap of unknown length
22200 24025: contig of 1826 bp in length
24026 24125: gap of unknown length
24126 25587: contig of 1462 bp in length
25588 25687: gap of unknown length
25688 27023: contig of 1336 bp in length
27024 27123: gap of unknown length
27124 28445: contig of 1522 bp in length
28446 28745: gap of unknown length
28746 30266: contig of 1881 bp in length
30267 30726: gap of unknown length
30727 32587: contig of 1861 bp in length
32588 32687: gap of unknown length
32688 34188: contig of 1501 bp in length
34189 34288: gap of unknown length
34289 36119: contig of 1831 bp in length
36120 36219: gap of unknown length
36220 38258: contig of 2039 bp in length
38259 38358: gap of unknown length
38359 40378: contig of 2520 bp in length
40379 40978: gap of unknown length
40979 42144: contig of 1166 bp in length
42145 42244: gap of unknown length
42245 44391: contig of 2047 bp in length
44392 44391: gap of unknown length
44392 46537: contig of 2146 bp in length
46538 46637: gap of unknown length
46638 48861: contig of 2224 bp in length
48862 48961: gap of unknown length
48962 52378: contig of 3417 bp in length
52379 52478: gap of unknown length
52479 56076: contig of 3598 bp in length
56077 56176: gap of unknown length
56177 58172: contig of 2576 bp in length
58173 58552: gap of unknown length
58553 62195: contig of 3343 bp in length
62196 62295: gap of unknown length
62296 65420: contig of 3125 bp in length
65421 68371: contig of 2851 bp in length
68372 68471: gap of unknown length
68472 70830: contig of 2359 bp in length
70831 70931: gap of unknown length
70932 74761: contig of 3831 bp in length
74762 74861: gap of unknown length
74862 77970: contig of 3109 bp in length
77971 78070: gap of unknown length
78071 79767: contig of 1697 bp in length
79768 79867: gap of unknown length
79868 82885: contig of 3018 bp in length
82886 82985: gap of unknown length
82986 87667: contig of 4682 bp in length
87668 87767: gap of unknown length
87768 90237: contig of 2470 bp in length
90238 90337: gap of unknown length
90339 93798: contig of 3461 bp in length
93799 93898: gap of unknown length
93899 97743: contig of 3845 bp in length
97744 97843: gap of unknown length
97844 99932: contig of 2089 bp in length
99933 100032: gap of unknown length
100033 104104: contig of 4072 bp in length
104105 104204: gap of unknown length
104205 108817: contig of 4613 bp in length
108818 108917: gap of unknown length
108919 114327: contig of 5410 bp in length
114328 114427: gap of unknown length
114428 118321: contig of 3894 bp in length
118322 118421: gap of unknown length
118422 122579: contig of 4158 bp in length
122580 122679: gap of unknown length

* 12680 126333: contig of 3654 bp in length
* 126334 126433: gap of unknown length
* 131306 131306: contig of 4873 bp in length
* 131307 131406: gap of unknown length
* 131407 137457: contig of 6051 bp in length
* 137458 137557: gap of unknown length
* 137558 141574: contig of 4017 bp in length

Alignment Scores:

Pred. No.: 3.06e+03 Length: 184024
Score: 89.00 Matches: 43
Percent Similarity: 37.34% Conservative: 16
Best Local Similarity: 27.22% Mismatches: 58
Query Match: 13.17% Indels: 41
DB: 2 Gaps: 6

US-10-048-196-2 (1-134) x AC127124 (1-184024)

Qy 5 LeuTyrValThrAlaThrLeuMetThrAla-----Phe 15
||| ||| ::|||::|||::|||::|||
Db 57327 CTTGGTGTCAATCCACACTGATTACGGCTGACGTTGATGCTGGGGCTTCCTATTATTT 57268
Qy 16 ThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSer-----Ser 32
||||| |||||::|||::|||::|||
Db 57267 ATTTGGCCATATTGGATCCACCAAGGGGTATCCATCTACTCAGCCCTCTTTTCT 57208
Qy 33 AlaAsnLeuThrSerLeuLe-----LysHisAlaValLysGlnThrCysGln--- 49
||| |||::|||::|||::|||::|||
Db 57207 GCTCCACTCAGTACTACATGGTCTCCAGCTCTCAGGAATAAATTACACCTCTCAATGC 57148
Qy 50 -----ThrGlnLeuThrGly-HisG1 56
||| |||
Db 57147 TTTAGAGCAAGTAGCTCCCAAGAGAGTCACCTCTGCACCTCTTTGATATCCATCC 57088
Qy 56 nTyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluTh 76
::: ::|||::|||::|||::|||
Db 57087 TAGCTTTCACTTCTCTGCTCTGTTGCTTGGGATAGGACGAGCTGCGAATC 57028
Qy 76 rAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrTh 96
||| |||||::|||::|||::|||
Db 57027 AGGCAAGTGGGCAAGAGAACAGAACCAACCAATCAGGACGGGCAAAATAGAAAC 56968
Qy 96 rAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerle 116
||| |||::|||::|||::|||
Db 56967 AGCA-----GTAAACACTGAG-----CACAAATT 56944
Qy 116 ulysProCysMetLeuGluThrValAsnAlaPheIleValProThrThrThr 133
||||| ||| |||||::|||::|||
Db 56943 AAAGCCTTGCCCACTGATGCTGCTTCTACCTTTCAATGTCCTGTCCACCACC 56892

RESULT 20
BD003708/c 5857 bp DNA linear PAT 31-JAN-2002
LOCUS BD003708
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003708
VERSION BD003708.1 GI:18631669
KEYWORDS JP 2001501833-A/28.
SOURCE unidentified.
ORGANISM Streptococcus pneumoniae

REFERENCE 1 (bases 1 to 5857)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C., Fannon,M. and Dougherty,B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 28 13-FEB-2001;
HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001501833-A/28
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH,GIL H CHOI,PATRICK J DILLON,CRAIG A ROSEN, PI STEVEN C BARASH,

PI MICHAEL FANNON,BRIAN A DOUGHERTY
PC C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
PC C12N1/21,
PC C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..5857 /organism='Unidentified'.
FEATURES
Source Location/Qualifiers
1..5857 /organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 1800 a 1173 c 814 g 2070 t
ORIGIN

Alignment Scores:

Pred. No.: 47.2 Length: 5857
Score: 88.50 Matches: 28
Percent Similarity: 48.54% Conservative: 22
Best Local Similarity: 27.18% Mismatches: 38
Query Match: 13.09% Indels: 15
DB: 6 Gaps: 4

US-10-048-196-2 (1-134) x BD003708 (1-5857)

Qy 30 AsnSerSerAlaAsnLeuThrThrSerLeuIleLysHis----- 42
||| ::|||::|||::|||::|||
Db 5016 RATGAATGATTAACCTTACGTCCTCAAGCAATTAATTCATAGATGCCATGTTACCGAG 4957
Qy 43 ---AlaValLysGlnThrCysGln-----ThrGlnLeuThrGlyHisGlnTyr 57
::: ::|||::|||::|||::|||
Db 4956 CCAAGCATAGACAAAGCATGTCAAAAGCAGGGGTATCAGCGCAACAGGTGATTAAGTAT 4897
Qy 58 TrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAla---LysIleSerGluThr 76
|||||::|||::|||::|||::|||
Db 4896 CTAAAGTTGCGCGGTATAAAGACATGAGGATAAGCAAGATGAGATGATGATAAA 4837
Qy 77 AlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGlu-----Leu 94
::: ::|||::|||::|||::|||
Db 4836 ACAACTCAATGCTTTATTAGCATCATCTACGCTGTTCTGTACTCAATCATATTATG 4777
Qy 95 ThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHis 114
::: ::|||::|||::|||::|||
Db 4776 ATGGACAGTAAAGTCAACCGTTTATAAGAACTCAAGCAAGGCTATCTTGAACAA 4717
Qy 115 SerLeuLys 117
||| |||
Db 4716 TCTTATAAG 4708

RESULT 21
AE007414 11849 bp DNA linear BCT 31-AUG-2001
LOCUS AE007414
DEFINITION Streptococcus pneumoniae TIGR4 section 97 of 194 of the complete genome.
ACCESSION AE007414 AE005672
VERSION AE007414.1 GI:14972604
KEYWORDS
SOURCE Streptococcus pneumoniae TIGR4.
ORGANISM Streptococcus pneumoniae TIGR4
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

REFERENCE 1 (bases 1 to 11849)
AUTHORS Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D., Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T., Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.


```
gene AC111203
CDS AC111203.5 GI:21716965
      HTG: HTGS-PHASE1.
      SOURCE: Norway rat.
      ORGANISM: Rattus norvegicus
      Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
      Rattus.
      1 (bases 1 to 227634)
      Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
      Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T.,
      Barbarella, J., Benton, J., Bimonte, K., Blankenburg, K., Bonnin, D.,
      Bouck, J., Bowles, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
      Buay, C., Burch, P., Burrell, C., Burrell, K.L., Byrd, N.C.,
      Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
      Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
      Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
      Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
      Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
      Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
      Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
      Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
      Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
      Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
      Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
      Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B.,
      Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
      Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
      Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
      Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
      Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisse, H.,
      Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
      Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
      Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
      Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
      Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
      Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okuwonu, G.,
      Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
      Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
      Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
      Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
      Sodergren, E., Sotak, T., Sparks, A., Stanley, N., Stone, H.,
      Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
      Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
      Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
      Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
      Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
      Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
      Weinstock, G., and Gibbs, R.
      Direct Submission
      Unpublished
      2 (bases 1 to 227634)
      Worley, K.C.
      Direct Submission
      Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
      of Molecular and Human Genetics, Baylor College of Medicine, One
      Baylor Plaza, Houston, TX 77030, USA
      3 (bases 1 to 227634)
      Worley, K.C.
      Direct Submission
      Submitted (09-JUL-2002) Human Genome Sequencing Center, Department
      of Molecular and Human Genetics, Baylor College of Medicine, One
      Baylor Plaza, Houston, TX 77030, USA
      On Jul 9, 2002 this sequence version replaced gi:20467446.
      ----- Genome Center
      Center: Baylor College of Medicine
      Center code: BCM
      Web site: http://www.hgsc.bcm.tmc.edu/
      Contact: hgsc-help@bcm.tmc.edu
      ----- Project Information
      -----
      Center project name: GCOJ
      Center clone name: CH230-7H1
```

```
AC111203
CDS AC111203.5 GI:21716965
      HTG: HTGS-PHASE1.
      SOURCE: Norway rat.
      ORGANISM: Rattus norvegicus
      Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
      Rattus.
      1 (bases 1 to 227634)
      Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
      Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T.,
      Barbarella, J., Benton, J., Bimonte, K., Blankenburg, K., Bonnin, D.,
      Bouck, J., Bowles, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
      Buay, C., Burch, P., Burrell, C., Burrell, K.L., Byrd, N.C.,
      Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
      Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
      Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
      Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
      Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
      Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
      Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
      Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
      Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
      Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
      Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
      Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B.,
      Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
      Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
      Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
      Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
      Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisse, H.,
      Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
      Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
      Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
      Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
      Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
      Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okuwonu, G.,
      Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
      Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
      Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
      Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
      Sodergren, E., Sotak, T., Sparks, A., Stanley, N., Stone, H.,
      Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
      Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
      Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
      Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
      Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
      Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
      Weinstock, G., and Gibbs, R.
      Direct Submission
      Unpublished
      2 (bases 1 to 227634)
      Worley, K.C.
      Direct Submission
      Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
      of Molecular and Human Genetics, Baylor College of Medicine, One
      Baylor Plaza, Houston, TX 77030, USA
      3 (bases 1 to 227634)
      Worley, K.C.
      Direct Submission
      Submitted (09-JUL-2002) Human Genome Sequencing Center, Department
      of Molecular and Human Genetics, Baylor College of Medicine, One
      Baylor Plaza, Houston, TX 77030, USA
      On Jul 9, 2002 this sequence version replaced gi:20467446.
      ----- Genome Center
      Center: Baylor College of Medicine
      Center code: BCM
      Web site: http://www.hgsc.bcm.tmc.edu/
      Contact: hgsc-help@bcm.tmc.edu
      ----- Project Information
      -----
      Center project name: GCOJ
      Center clone name: CH230-7H1
```

RESULT 22

AC111203

LOCUS

DEFINITION Rattus norvegicus clone CH230-7H1, *** SEQUENCING IN PROGRESS ***

----- Summary Statistics

Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 175895 bases at least Q40
Consensus quality: 182735 bases at least Q30
Consensus quality: 188082 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently
consists of 64 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1
1227: contig of 1227 bp in length
1228
1327: gap of unknown length
1328
1328: contig of 1504 bp in length
2832
2831: gap of unknown length
2832
2832: contig of 1263 bp in length
2832
4194: contig of 1263 bp in length
4195
4234: gap of unknown length
5804: contig of 1510 bp in length
5805
5904: gap of unknown length
5905
6963: contig of 1059 bp in length
6964
7063: gap of unknown length
7064
8514: contig of 1451 bp in length
8515
8614: gap of unknown length
8615
9061: contig of 1347 bp in length
9062
10061: gap of unknown length
10062
11183: contig of 1122 bp in length
11184
12384: contig of 1101 bp in length
12385
12484: gap of unknown length
12485
13693: contig of 1209 bp in length
13694
13793: gap of unknown length
13794
15094: contig of 1301 bp in length
15095
15194: gap of unknown length
15195
16828: contig of 1634 bp in length
16829
16928: gap of unknown length
16929
18450: contig of 1522 bp in length
18451
18550: gap of unknown length
18551
20556: contig of 2006 bp in length
20557
20656: gap of unknown length
20657
21826: contig of 1170 bp in length
21827
21926: gap of unknown length
21927
23400: contig of 1474 bp in length
23401
23500: gap of unknown length
23501
24836: contig of 1336 bp in length
24837
24936: gap of unknown length
24937
26065: contig of 1129 bp in length
26066
26165: gap of unknown length
26166
28322: contig of 2157 bp in length
28323
28422: gap of unknown length
30029: contig of 1607 bp in length
30030
30129: gap of unknown length
30130
31768: contig of 1639 bp in length
31769
31868: gap of unknown length
33450: contig of 1582 bp in length
33451
33550: gap of unknown length
33551
35137: contig of 1587 bp in length
35138
35237: gap of unknown length
35238
37307: contig of 2070 bp in length
37308
37407: gap of unknown length
39940: contig of 2533 bp in length
39941
40040: gap of unknown length
40041
42352: contig of 2312 bp in length
42353
42452: gap of unknown length
42453
44486: contig of 2034 bp in length
44487
44586: gap of unknown length
44587
46934: contig of 2368 bp in length

46955
47054: gap of unknown length
47055
48656: contig of 1602 bp in length
48657
48756: gap of unknown length
48757
50781: contig of 2025 bp in length
50782
50881: gap of unknown length
50882
53948: contig of 3066 bp in length
54047: gap of unknown length
54048
56742: contig of 2695 bp in length
56743
56842: gap of unknown length
56843
60070: contig of 3228 bp in length
60071
63329: contig of 3159 bp in length
63330
63429: gap of unknown length
63430
64980: contig of 1551 bp in length
64981
65080: gap of unknown length
65081
68309: contig of 3229 bp in length
68310
68409: gap of unknown length
68410
70974: contig of 2565 bp in length
70975
71074: gap of unknown length
71075
75469: contig of 4395 bp in length
75470
75569: gap of unknown length
75570
78406: contig of 2837 bp in length
78407
78506: gap of unknown length
81000: contig of 2494 bp in length
81001
81100: gap of unknown length
81101
84692: contig of 3592 bp in length
84693
84792: gap of unknown length
84793
88894: contig of 4102 bp in length
88895
88994: gap of unknown length
88995
93304: contig of 4310 bp in length
93305
93404: gap of unknown length
93405
96551: contig of 3147 bp in length
96552
102146: contig of 5495 bp in length
102147
102246: gap of unknown length
102247
105682: contig of 3416 bp in length
105683
105762: gap of unknown length
105763
110846: contig of 5084 bp in length
110847
110946: gap of unknown length
110947
116120: contig of 5174 bp in length
116121
116220: gap of unknown length
116221
120740: contig of 4520 bp in length
120741
120840: gap of unknown length
120841
126090: contig of 5250 bp in length
126091
126190: gap of unknown length
126191
131454: contig of 5264 bp in length
131455
131554: gap of unknown length
131555
136229: contig of 4675 bp in length
136230
136329: gap of unknown length
136330
139840: contig of 3511 bp in length
139841
146131: contig of 6191 bp in length

Alignment Scores:

Pred. No.: 4.51e+03 Length: 227634
Score: 88.50 Matches: 36
Percent Similarity: 44.35% Conservatives: 19
Best Local Similarity: 29.03% Mismatches: 46
Query Match: 13.09% Indels: 24
DB: 2 Gaps: 5

US-10-048-196-2 (1-134) x AC111203 (1-227634)

QY 11 LeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsn 30
DB 153588 CTTCTGCCGCCCTTCTCTGTCCTTACCCCTCTCTCTCCAGAGCAACTGTAGAAAC 153647
QY 31 SerSerAlaAsnLeuThrThr-----SerLeuIleLysHisAlaValLysGln 46
DB 153648 CAAAATTTCTTCATTTTCAACATGGCCCATAGGCTCGATTATT-----GCCATTTCTGCAA 153701
QY 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLysLeuSer 66

```

Db 153702 -----ACTGGCCACAGAACCTGGAGCATTGCACTGCAAGCATTTCC 153743
Qy 67 SerGlu-----SerLysAlaLysIleSerGluThrAlaCysGlyCysValAla 82
Db 153744 TCCACCTTTCTGTGTGAGAACCTGCCAGGAACAGTGTCTATCTCTGTCATGTC 153803
Qy 83 AspLysAla-----ProGluAlaValSerLeuThrGluLeuThr 95
Db 153804 CAGAAAGCACCTGGGCCCTGTTCAGGAGGAGCAGAACAGCAGCTAGGAGTCTACA 153863
Qy 96 ThrAlaAlaIleAsnProAsnAlaArgThrGluAlaLalaGlnLysIleValArgHisSer 115
Db 153864 AAGCCTTGCTAACCCCTCTCTGCTA-CCCATGTGCCCAAGTCTAGCAGTAGATGGAGC 153922
Qy 116 LeuLysProCys 119
Db 153923 CTTCATAGATCT 153934

RESULT 23
103335
LOCUS I03335 2000 bp ss-DNA linear PAT 21-MAY-1993
DEFINITION Sequence 2 from Patent US 4826765.
ACCESSION I03335
VERSION I03335.1 GI:270041
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2000)
AUTHORS Greene,F.C., Stiles,J.I., Neill,J.D., Anderson,O.D. and Litts,J.C.
TITLE Yeast strains genetically engineered to produce wheat gluten
JOURNAL Patent: US 4826765-A 2 02-MAY-1989;
The United States of America as represented by the Secretary of
the; Washington, DC
FEATURES
source
location/Qualifiers
1..2000
/organism="unknown"
BASE COUNT 686 a 481 c 352 g 480 t 1 others
ORIGIN

Alignment Scores:
Pred No.: 14 Length: 2000
Score: 88.00 Matches: 36
Percent Similarity: 37.50% Conservative: 18
Best Local Similarity: 25.00% Mismatches: 52
Query Match: 13.02% Indels: 38
DB: 6 Gaps: 6

US-10-048-196-2 (1-134) x I03335 (1-2000)
Qy 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
Db 1113 GTATTTCGCAACCAACAACCAATTTACAGCAGCAGCAGCAGCAGCAGCAACA 1172
Qy 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
Db 1173 ACAACAACAACAACAACAACAATCTCTCAACAATTTTGCAACAACAACCTGATTCATG 1232
Qy 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrrTrpLys 59
Db 1233 -----CATGGATGTTGTATTGCGAACAACAACAACATAGCGCATGGAAG 1274
Qy 60 Ile-----AlaAlaMetLysLysSerSerGluSerLysAlaLysIleSer----- 74
Db 1275 ATCACAAGTTTGGACAAAGTACTTACCAGCTGTGCAAGAAATGTGTGTCACACCT 1334
Qy 75 -----GluThrAlaCysGlyCysValAla 82
Db 1335 ATGCAGATCCCTGAGCAGTCGACAGTGCAGGCCATCCTCAAGTGTGTTTCATGCTAT--- 1391
Qy 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102

```

```

Db 1392 -----TATTTCGATCAACAACAACAACAACAACAACCATCGCCAGGT 1439
Qy 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
Db 1440 CTCCTTCCAAAGCCTCTGCAACAATA-----TCCATTAGCCAGGGCTCCTTCG 1490
Qy 123 ThrValAsnAla 126
Db 1491 GCCATCTCAGCA 1502

RESULT 24
TAGLIAA
LOCUS TAGLIAA 2347 bp DNA linear PLN 12-SEP-1993
DEFINITION Wheat gene for storage protein gliadin (A-gliadin subgroup).
ACCESSION X01130
VERSION X01130.1 GI:21754
KEYWORDS gliadin; inverted repeat; storage protein.
SOURCE Triticum aestivum.
ORGANISM Triticum aestivum
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 2347)
AUTHORS Anderson,O.D., Litts,J.C., Gautier,M.F. and Greene,F.C.
TITLE Nucleic acid sequence and chromosome assignment of a wheat storage
protein gene
JOURNAL Nucleic Acids Res. 12 (21), 8129-8144 (1984)
MEDLINE 85062803
PUBMED 6095191
FEATURES
source
location/Qualifiers
1..2347
/organism="Triticum aestivum"
/db_xref="taxon:4565"
53..61
/note="inverted repeat A"
59..102
/note="imp. inverted repeat B"
62..70
/note="inverted repeat A'"
76..83
/note="pot. TATA-box"
82..89
/note="inverted repeat C"
90..97
/note="inverted repeat C'"
631..641
/note="pot. CAAT elements"
668..675
/note="pot. TATA-box"
713..724
/note="imp. inverted repeat D"
772..1632
/note="storage protein (aa 1-286)"
/codon_start=1
/protein_id="CAA25593.1"
/db_xref="GI:21755"
/translacion="MKTFLILVLAIVATTATTAVRPVPOLOPQNPSQQOPOEQVPL
VQQQFLGQQPFPQQPYPOPPFPOLYLQLOPFPOLPYSPQPPRQPPYPO
PQPSIQPPIQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
SQVLOQSTYQLLQELCCOHLWQIPEQSOCAILKLVHAILHQQKQKQSSQV
QPLQQYPLGGSGFRPSQNPQAGSVQPPQLPQPEIRNLALQTLPMNCNVIP
IAPFGIGFTN"
1623..1635
/note="imp. inverted repeat D'"
1640..1650
/note="imp. direct repeat E"
1651..1661
/note="imp. direct repeat E'"
1710..1715
/note="pot. polyadenylation signal"
1767..1772
/note="pot. polyadenylation signal"

```



```

repeat_unit 1793..1828
misc_feature /note="imp. inverted repeat B'"
repeat_unit /note="polyA site in cDNA clones"
repeat_unit /note="inverted repeat F"
repeat_unit /note="inverted repeat F'"
repeat_unit /note="inverted repeat G"
repeat_unit /note="inverted repeat G'"

BASE COUNT 821 a 546 c 385 g 595 t
ORIGIN

Alignment Scores:
Pred. No.: 17 Length: 2347
Score: 88.00 Matches: 36
Percent Similarity: 37.50% Conservative: 18
Best Local Similarity: 25.00% Mismatches: 52
Query Match: 13.02% Indels: 38
DB: 8 Gaps: 6

US-10-048-196-2 (1-134) x TAGLIAA (1-2347)
QY 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
Db 1089 GTATTGCGAACCAACCAACCAATTTTCACGAGCAGCAGCAGCAGCAGCAGCAACA 1148

QY 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
Db 1149 ACAACAACAACAACAACAACAATCTTCAACAAATTTTGCACCAACAACATGATCCATG 1208

QY 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLys 59
Db 1209 -----CATGGATGTTGATTGTCAGCAACAACAACAATAGCCATGGAG 1250

QY 60 Ile-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer----- 74
Db 1251 ATCAACAAGTTTGGCAACAAGTACTTACCAGCTGTTGCAAGAATTGTGTGCAACACT 1310

QY 75 -----GluThrAlaCysGlyCysValAla 82
Db 1311 ATGGCAGATCCCTGAGCAGTCCGAGTGCAGGCCATCTCTCAAGTGTTCATGCTAT--- 1367

QY 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102
Db 1368 -----TATTCGTCATCAACAACAACAACAACAACAACAACCAATCCAGCCAGGT 1415

QY 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
Db 1416 CTCCTTCCAACAGCCTCTGCAACAATA-----TCCATTAGGCCAGGGCTCTTCGG 1466

QY 123 ThrValAsnAla 126
Db 1467 GCCATCTCAGCA 1478

RESULT 25
CNS08CAU/c CNS08CAU 157362 bp DNA linear HTG 09-AUG-2002
LOCUS Oriza sativa chromosome 12 clone OSJNBa0085B23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 5 ordered pieces.
ACCESSION AL844497
VERSION AL844497.2 GI:22208182
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEPIN.
SOURCE Oriza sativa
ORGANISM Oriza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoideae; Oryza.
REFERENCE 1 (bases 1 to 157362)
AUTHORS Choinsne,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
Segurens,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,

```

Weissenbach,J. and Quetier,F.
 Oriza sativa chromosome 12 sequencing
 2 (bases 1 to 157362)
 Genoscope.
 Direct Submission
 Submitted (02-AUG-2002) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 On Aug 11, 2002 this sequence version replaced gi:21953218.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence.
 work on the sequence is in progress and the release of this data is
 based on the understanding that the sequence may change as work
 continue. The sequence may be contaminated with foreign sequence
 from E.coli, yeast, vector, phage, etc.
 Contigs composition :
 31145 bp contig from 1 to 31145
 46320 bp contig from 31246 to 77565
 7663 bp contig from 77666 to 85328
 14726 bp contig from 85429 to 100154
 57108 bp contig from 100255 to 157362.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1. 31145: contig of 31145 bp in length
 * 31146 31245: gap of 100 bp
 * 31246 77565: contig of 46320 bp in length
 * 77566 77665: gap of 100 bp
 * 77666 85328: contig of 7663 bp in length
 * 85329 85428: gap of 100 bp
 * 85429 100154: contig of 14726 bp in length
 * 100155 100254: gap of 100 bp
 * 100255 157362: contig of 57108 bp in length.
 FEATURES
 Location/Qualifiers
 1..157362
 /organism="Oriza sativa"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="12"
 /clone="OSJNBa0085B23"
 /clone_lib="OSJNBa"
 BASE COUNT 45547 a 32463 c 32507 g 46443 t 402 others
 ORIGIN

Alignment Scores:
 Pred. No.: 3, 21e+03 Length: 157362
 Score: 88.00 Matches: 35
 Percent Similarity: 46.96% Conservative: 19
 Best Local Similarity: 30.43% Mismatches: 54
 Query Match: 13.02% Indels: 7
 DB: 2 Gaps: 3
 US-10-048-196-2 (1-134) x CNS08CAU (1-157362)
 QY 19 SerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSer 38
 Db 41529 ACTATTGCATTTCTTCCT---AGCTCTCCCAAGAGCTCTTTTCAGTCGCTCTCTCAAC 41473
 QY 39 LeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTr 58
 Db 41472 ATGATAGCGATTTCATGTCACCAACAGTTTGTATCATCGAACTCCACACTACATTTCTT 41413
 QY 58 physile-----AlaAlaMetLysLeuSerGluSerLysAlaLysIleSerG 75
 Db 41412 GATAACAGAAATTAAGGCATGTGCGCGCTTTCTTCCGCATCATCATGTCTATAGCTGA 41353

LOCUS	TAGLIA	3022 bp	mRNA	linear	PLN 12-SEP-1993
TAGLIA					

QY 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
 Db 973 ACAACAACAACAACAACAATCTTCAACAATAATTTGCAACAACAACATGATTCATG 1032
 QY 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLys 59
 Db 1033 -----CATGGATGTTGTTATTCGAGCAACAACAACATAGCGCATGGAAG 1074
 QY 60 Ile-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer----- 74
 Db 1075 ATCACAAGTTTTCACAACAAGTACTTACCAGCTGTTGCAGAATTTGTTGTCACACCT 1134
 QY 75 -----TATTCGTCATCAACAACAACAACAACAACAACAACAACATCGACCGAGT 1191
 Db 1135 ATGCAGATCCCTGAGCAGTCGCGAGTCGCCAGGCGATCCCAATGTTGTCATGCTAT--- 1191
 QY 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102
 Db 1192 -----TATTCGTCATCAACAACAACAACAACAACAACAACAACATCGACCGAGT 1239
 QY 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
 Db 1240 CTCCTTCCACAGCCTCTGCAACAATA-----TCCATTAGCGAGGCTCCTTCCG 1290
 QY 123 ThrValAsnAla 126
 Db 1291 GCCATCTCAGCA 1302
 RESULT 31
 LOCUS WHTGLIABE 3022 bp mRNA linear PLN 27-APR-1993
 DEFINITION Wheat ('T.aestivum) alpha/beta gliadin class I gene, clone pW8233, complete cds.
 ACCESSION K03076.1 GI:170719
 VERSION 1
 KEYWORDS gliadin.
 SOURCE Wheat ('T.aestivum) cv. Yamhill) endosperm, cDNA to mRNA, clone pW8233, complete cds.
 ORGANISM
 Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.
 1 (sites)
 Summer-Smith, M., Rafalski, J.A., Sugiyama, T., Stoll, M. and Soell, D. Conservation and variability of wheat alpha/beta-gliadin genes Nucleic Acids Res. 11, 3905-3916 (1983)
 2 (bases 1 to 3022)
 Rafalski, J.A., Scheets, K., Metzler, M., Peterson, D.M., Hedgcock, C. and Soell, D.G.
 Developmentally regulated plant genes: the nucleotide sequence of a wheat gliadin genomic clone EMBO J. 3 (6), 1409-1415 (1984)
 84261434
 6204862
 3 (bases 1 to 3022)
 Soell, D.
 Unpublished (1985)
 [3] revision.
 [1] sites; mRNA start, p-boxes.
 Sequence in computer-readable form provided by D. Soell, 04-JUL-1985.
 Gliadins, the major wheat seed storage proteins, are encoded by a multigene family and are highly conserved. In the sequence presented here a TATA box is located at positions 492-499 and potential poly-A signals at 1534-1539 and 1591-1596. Four P-boxes (containing many proline codons) are found at 758-793, 794-829, 830-868, and 869-904. The P-box presumably arose in the ancestral alpha/beta-gliadin gene and was multiplied prior to the extensive multiplication of the whole gene. Every P-box is slightly different from every other. Two polyglutamine stretches reside at 941-994 and 1202-1224. In wheat DNA 'cg' and 'cng' sequences are over 80% methylated to m5C. An alternating purine pyrimidine

stretch with 2-forming potential lies between nucleotides 2258 and 2281.

FEATURES
 Source Location/Qualifiers
 1..3022
 /organism="Triticum aestivum"
 /db_xref="taxon:4565"
 529..1626
 /product="gliadin mRNA"
 596..1456
 /note="alpha/beta-gliadin precursor"
 /codon_start=1
 /protein_id="AA34280.1"
 /db_xref="GI:170720"
 596..655
 /note="alpha/beta-gliadin signal peptide"
 656..1453
 /product="alpha/beta-gliadin"
 BASE COUNT 1021 a 741 c 521 g 739 t
 ORIGIN 1 bp upstream of HindIII site.
 Alignment Scores:
 Pred. No.: 29.7 Length: 3022
 Score: 87.00 Matches: 36
 Percent Similarity: 36.81% Conservative: 17
 Best Local Similarity: 25.00% Mismatches: 53
 Query Match: 12.87% Indels: 38
 DB: 8 Gaps: 6
 US-10-048-196-2 (1-134) x WHTGLIABE (1-3022)
 QY 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
 Db 913 GTATTGCAACCAACAACAACCAATTTTCACAGCAGCAGCAGCAGCAGCAACA 972
 QY 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
 Db 973 ACAACAACAACAACAACAACCAATCTTCAACAATAATTTGCAACAACAACATGATTCATG 1032
 QY 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLys 59
 Db 1033 -----CATGGATGTTGTTATTCGAGCAACAACAACATAGCGCATGGAAG 1074
 QY 60 Ile-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer----- 74
 Db 1075 ATCACAAGTTTTCACAACAAGTACTTACCAGCTGTTGCAGAATTTGTTGTCACACCT 1134
 QY 75 -----TATTCGTCATCAACAACAACAACAACAACAACAACAACATCGACCGAGT 1191
 Db 1135 ATGCAGATCCCTGAGCAGTCGCGAGTCGCCAGGCGATCCCAATGTTGTCATGCTAT--- 1191
 QY 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102
 Db 1192 -----TATTCGTCATCAACAACAACAACAACAACAACAACAACATCGACCGAGT 1239
 QY 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
 Db 1240 CTCCTTCCACAGCCTCTGCAACAATA-----TCCATTAGCGAGGCTCCTTCCG 1290
 QY 123 ThrValAsnAla 126
 Db 1291 GCCATCTCAGCA 1302
 RESULT 32
 LOCUS AP003251/c 149145 bp DNA linear
 DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC clone: P0446505.

ACCESSION
VERSION
KEYWORDS
SOURCE

AF003251
AP003251.3 GI:20160607

Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:P0446B05
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
AUTHORS
TITLE

1
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0446B05
Published Only in Database (2001)

REFERENCE
AUTHORS
TITLE

2
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission

JOURNAL
TITLE

Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@ias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT

On Apr 16, 2002 this sequence version replaced gi:17933036.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI NonRedundant protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologs of the coding regions were searched against
NCBI NonRedundant protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from SP6 to T7 of the PAC clone.
This sequence of P0446B05 clone has an overlap with P0406G08
(DBJ:AP003240) clone at the position 1 to 17,096 of 5' end and
with P0005H10 (DBJ:AP004127) at the position 98,039 to 149,145 of
3' end. Detailed information on overlap and assembly quality
together with annotation of this entry is available at
http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES
source

1. .149145
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clones="P0446B05"
complement(join(1108..1218,2391..2486,2568..2681,
2756..2833,2914..3015,4386..4661))
/gene="P0446B05.1"
complement(join(1108..1218,2391..2486,2568..2681,
2756..2833,2914..3015,4386..4661))
/gene="P0446B05.1"
/note="contains ESTs D23149(C2313),C26277(C12019)
similar to Arabidopsis thaliana chromosome 3,At3g09320
unknown protein"
/codon_start=1
/protein_id="BAB89554.1"
/db_xref="GI:20160608"
/translation="MAGRCGYVTVPILSLAAATCYVYVAFELAPWLGLATAAGVA
NAVATLAAACVATYAVSRDGRVPPAFPLDVEDAESPIEIKRKHCIWNNC
VGHENYKIFLVYVAVSLVIGGAVSLPKNEQLGSSSTSVSNVQIIC
GVFCPLALALSILLGHVYLIIFHNKTTIEHGVRAWLAERAGNLYHHPYDLGVYE
NLVSLGPNALCMWLCPIRSRTGIRERTSYDIPLSLPTPI"
complement(11360..12688)
/gene="P0446B05.2"
complement(11360..12688)

gene

CDS

gene

CDS

/gene="P0446B05.2"
/note="contains EST AU176487(E21047)
similar to Arabidopsis thaliana chromosome 5, At5g02190
unknown protein"
/codon_start=1
/protein_id="BAB89555.1"
/db_xref="GI:20160609"
/translation="MPPPALPVCLLLLLLLLVPRPAAAAAATRLPLFLRLAR
QVPAGALPRPASKRLTFASVPCDSAOCSRDLPPACDAGSKOCRVSLSVADGSSGCG
GGRSALSRPRASLTFTASVPCDSAOCSRDLPPACDAGSKOCRVSLSVADGSSGCG
ALATEFTVGQGPLRAAFCGMATFDTSPGVATAGLLCMNRGALSVSQALSTRRES
YCSISDRDAGVLLHSHDLFLFLNVTPLXQAMPPLPYFDRVAYSVOLLLRIGVKKPL
PTPASVLAPDHTGAGTMDVDSGTQFTLLGDAYSALUKAESRQTKPWLPAINDPNAF
QEADPTCFRYPQGRAPPARLPAPVTLFNGAQMVTAGDRLLYKYPGERRGDGVNCLTF
GNADMVPTAYIVICHHNMNVWEYDLERGVGLAIRCDDVASERLGLML"
join(13981..14067,14601..14684,14736..14909)
/gene="P0446B05.3"
join(13981..14067,14601..14684,14736..14909)
/gene="P0446B05.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB89556.1"
/db_xref="GI:20160610"
/translation="MGEKCIAMNAGSQARHTRLPDVTETSKDIIIVQDYTTTGS
LVDDKQSHCRDRVPPMPYSPGLHDLFLGDTIILDDDEHNSKRMSNRMRNDRR
LTALITADVTSN"
complement(17162..17869)
/gene="P0446B05.4"
complement(17162..17869)
/gene="P0446B05.4"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB89557.1"
/db_xref="GI:20160611"
/translation="MATFTAGTGMDSPIWCMCSRLHRPDLSTCTPRAPRAALE
EIVEVMDAGFLOACALRAPAAVSSSTQQLPTVTRDAGRTCAVCLDLPBGGA
VWPCDHAYHPQCIAPWLEAHDTCPLCRRESGLQVVEYQVQDGMVLSLSDGLVCL
MMPGKSEYRLGRVGRIFAVRVVDGTGLVKGVLRRUGSACHRFAPAAAGNLLSR
YRDCVIPNDLLLGVCQ"
join(22593..23216,24099..25639,26226..26367,26499..26606,
26691..26903,26990..27100,27933..28127,28751..28925,
29066..29103)
/gene="P0446B05.5"
join(22593..23216,24099..25639,26226..26367,26499..26606,
26691..26903,26990..27100,27933..28127,28751..28925,
29066..29103)
/gene="P0446B05.5"
/codon_start=1
/product="putative pumilio/Mpt5 family RNA-binding
protein"
/protein_id="BAB89558.1"
/db_xref="GI:20160612"
/translation="MGGGGGGDEGELELDLALGTGGAGARRRGDAGERELNLM
FRSGAPPTTEGSLNATISGLRGGEAAVVAIPDAETLNGHGLLSEDELRADPAY
LSYVSHGNLNPRLPPVLSKEDWRSTQRLKAGVVGIGDGRKVSPEETGHEPTVGRP
VFQNVGDFDEEARNVDGGAENVDGGGLGLSLGRQSRFADILQDNIGRTPTAS
EHPSVAVRSNFDLNOELLNSADQYSMHNDILEAQRAVGNQVNSGLPISNASQTP
FASVSGSISRNATPDSHYVPVPSGLPPGVVIRINSKELKNCSSSFNTASSKAVEP
DDTLAIANNMMSKDGTLGDSNYSQSKFQSEISDHQNISLDPKALOVNKNQHSML
ADTYLIGIPPTISQSPNPSFSDINKNVSLANRSTNTRIDGHAEMORSSTLSTRYQ
KSTPSSNASPGGSPAQHQNDINSAFLNYGLGYPPLSPGLPSMMNMCMGSGNNPPL
ESAAAASIAAFGADSRNLGNLILSSLSLSDVHNLRGTQATGLQQLPSLSPFFY
OYLKATQYAAQGTANCDSPSLERFGMSGQYGLTAVQAKYIEALLQOQKQYQMPGLGS
TTSNHGYGNLAFAGMGMAVPGSPGLSPVSPGSPSLRSLRSENRMRPSNLRNLGGN
SDPSGWNDFNPSSLLDEFKSNARSFELAEIASHVYFSDAQYGSRFIOOKLETATV
EEDMWFKETLMPQALSMTDVGNYVVKQFPEHGSAAQRRELAQDLGHVLSLQY
GCVRIOAKIEVWLDQTKVTLDELGHIMRCVRQDNQNHVIOKICEQVPDSIOFIIS
TFYGVVPLSTHPYGCVRIOVLEHCTDPKTYQDEIVMDLEILQSCVCMCAQDQNTVOH
VLEHGKPHERSVIIIEKLAGIQMSQOKFASNVYKCLTFGGPAERVLINEMIGTTD
ENEPLOQAMKMDQFCNYVYQVKVLETCDQDQRELILSRVKVHLSALKYTKGHIVARVE
KLVAAGERRITGLQAYPS"
30568..31146
/gene="P0446B05.6"

[illegible]

[illegible]

Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, P., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lounsbury, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Mines, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwu, S., Ogih, M., Okunishi, G., Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. D., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 185108)
Worley, K. C.

Direct Submission
Submitted (21-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 185108)
Worley, K. C.

Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20330712.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVTY
Center clone name: CH230-262L23
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 127499 bases at least Q40
Consensus quality: 136160 bases at least Q30
Consensus quality: 141175 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1022: contig of 1022 bp in length
* 1023 1122: gap of unknown length
* 1123 2137: contig of 1015 bp in length
* 2138 2237: gap of unknown length
* 2238 3423: contig of 1186 bp in length
* 3424 4780: contig of 1257 bp in length
* 4781 4880: gap of unknown length
* 4881 6070: contig of 1190 bp in length
* 6071 6170: gap of unknown length
* 6171 7206: contig of 1036 bp in length
* 7207 7306: gap of unknown length
* 7307 8639: contig of 1333 bp in length
* 8640 8739: gap of unknown length
* 8740 10239: contig of 1500 bp in length
* 10240 10339: gap of unknown length
* 10340 11596: contig of 1257 bp in length
* 11597 11696: gap of unknown length
* 11697 13560: contig of 1864 bp in length
* 13561 13660: gap of unknown length
* 13661 14957: contig of 1297 bp in length
* 14958 15057: gap of unknown length
* 15058 16247: contig of 1189 bp in length
* 16248 16346: gap of unknown length
* 16347 18173: contig of 1827 bp in length
* 18174 18273: gap of unknown length
* 18274 19773: contig of 1500 bp in length
* 19774 19873: gap of unknown length
* 19874 20937: contig of 1064 bp in length
* 20938 21037: gap of unknown length
* 21039 22694: contig of 1657 bp in length
* 22695 22794: gap of unknown length
* 22795 24383: contig of 1589 bp in length
* 24384 24484: gap of unknown length
* 24485 25704: contig of 1221 bp in length
* 25705 25804: gap of unknown length
* 25805 28163: contig of 2359 bp in length
* 28164 28263: gap of unknown length
* 28264 29401: contig of 1138 bp in length
* 29402 29501: gap of unknown length
* 29502 31633: contig of 2132 bp in length
* 31634 31734: gap of unknown length
* 31735 33921: contig of 2188 bp in length
* 33922 34021: gap of unknown length
* 34022 35757: contig of 1736 bp in length
* 35758 35857: gap of unknown length
* 35858 37514: contig of 1657 bp in length
* 37515 37614: gap of unknown length
* 37615 39730: contig of 2116 bp in length
* 39731 39830: gap of unknown length
* 39831 41739: contig of 1909 bp in length
* 41740 41839: gap of unknown length
* 41840 43994: contig of 2155 bp in length
* 43995 44094: gap of unknown length
* 44095 45581: contig of 1487 bp in length
* 45582 45681: gap of unknown length
* 45682 47275: contig of 1594 bp in length
* 47276 47375: gap of unknown length
* 47376 50938: contig of 3563 bp in length
* 50939 51038: gap of unknown length
* 51039 53405: contig of 2367 bp in length
* 53406 53505: gap of unknown length
* 53506 53885: contig of 1880 bp in length
* 53886 55485: gap of unknown length
* 55486 57173: contig of 1688 bp in length
* 57174 57273: gap of unknown length
* 57274 58840: contig of 1567 bp in length
* 58841 58940: gap of unknown length
* 58941 61253: contig of 2313 bp in length
* 61254 61354: gap of unknown length
* 61355 63243: contig of 1890 bp in length
* 63244 63443: gap of unknown length
* 63444 65848: contig of 2505 bp in length
* 65849 65948: gap of unknown length
* 65949 70327: contig of 4379 bp in length
* 70328 70427: gap of unknown length
* 70428 72825: contig of 2398 bp in length
* 72826 72925: gap of unknown length
* 72926 74961: contig of 2036 bp in length
* 74962 75061: gap of unknown length

Qy 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
Db 600 CTCCTTCCACAGCCTCAGCAACAATA-----TCCATCAGGCCAGCGCTTTTCCCA 650
Qy 123 ThrValAsnAla 126
Db 651 ACCATTTTCAGCA 662

RESULT 37
AR204236/c AR204236 3823 bp DNA llinear PAT 20-JUN-2002
DEFINITION Sequence 132 from patent US 6365723.
ACCESSION AR204236
VERSION AR204236.1 GI:21500834
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3823)
AUTHORS Blattner,F.R., Burland,V., Perna,N.T., Plunkett,G. and Welch,R.
TITLE Sequences of E. coli O157
JOURNAL Patent: US 6365723-A 132 02-APR-2002;
FEATURES Location/Qualifiers
source 1..3823
/organism="unknown"

BASE COUNT 1054 a 961 c 900 g 908 t
ORIGIN

Alignment Scores:
Pred. No.: 57.1 Length: 3823
Score: 85.50 Matches: 33
Percent Similarity: 36.91% Conservative: 22
Best Local Similarity: 22.15% Mismatches: 65
Query Match: 12.65% Indels: 29
DB: 6 Gaps: 3

US-10-048-196-2 (1-134) x AR204236 (1-3823)

Qy 6 TyrValThrAlaThrLeuMetThrAlaPhe---ThrLeuAlaSerCysAlaSerThrPro 24
Db 1462 TGGGTGACAGGTAAGTAGCGGTGAGTTTGGTAACTCTGGAGTACATCAGAACAA 1403
Qy 25 GluSerAsnProLysAsnSerSerAlaAsnLeuThrSerLeuIleLysHisAlaVal 44
Db 1402 GAAGGAAGTGGAGCACTGGAAGATCAACACAGCGCTATTCTTTCTCTTTGGTC 1343
Qy 45 LysGlnThrCysGlnThrGlnLeuThrGlnLysGlnTyrTrpLysIleAlaAlaMetLys 64
Db 1342 ATCCAGGCGCGGAACAGCAGCTCCGGCAAC----- 1310
Qy 65 LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLys 84
Db 1309 ATATCCACAGTGGCATATTCATCTTCAGCGACTGACTCGGCGAGCATTCAGATAAG 1250
Qy 85 AlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArg 104
Db 1249 ATGCTGCGCGTTTAGCATTAACCCAGGAATGAGCTGCGCATCTAACTCTGGCGTGCG 1190
Qy 105 ThrGluValAlaGlnLysIleValArgHisSerLeuLysProCys----- 119
Db 1189 GTAATCAGTTTGACAAGCGCTTTAGCCCACTCTTTATCTTCATGTTATTCAGCAAGACTA 1130
Qy 120 -----MetLeuGluThrValAsn 125
Db 1129 GCATAAGCCGCAATAGCCGCGGATGCTTTTACTGGCGCTTGCCTTAAGTATTCAGG 1070
Qy 126 AlaPheIleValProThrThrArg 134
Db 1069 TGCCTTTTCTTACCAATTACACCGAGG 1043

RESULT 38
AE005438/c

LOCUS AE005438 11010 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 57 of 290.
ACCESSION AE005438 AE005174
VERSION AE005438.1 GI:12516330
KEYWORDS
SOURCE Escherichia coli O157:H7 EDL933.
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 11010)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 11010)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES Location/Qualifiers
source 1..11010
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
101..829
/gene="molR_A"
/note="23283"
101..829
/gene="molR_A"
/function="putative regulator; Biosynthesis of cofactors, carriers: Molybdopterin"
/note="Residues 1 to 218 of 242 are 37.75 pct identical to residues 1 to 244 of 274 from Escherichia coli K-12 Strain MG1655: B2115"
/codon_start=1
/transl_table=11
/product="putative regulator (fragment)"
/protein_id="AAG57176.1"
/db_xref="GI:12516331"
/translation="MRHFLYODEKSHKFWAEQOQNELHINMGKVTNGOSQIKSFAD VAAEKAKLALIAETKKGIVENNASNVHIPPTKATPEVTSFESKNORPLWADAV IPVDDINRFAPHRSPREINLRKDGIEWRIADTRAYDPNNRYSYENWQQAF AEQRLDNNQQTGSAQSDAALLSWFSNSADELVDDLVIRCGLSEAVEIVLFAQL KYKPVKGAVTTIIPNHEAESLPS"
298..5428
/note="O-island #90: Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
1019..3910
/gene="molR_B"
/note="23284"
1019..3910
/gene="molR_B"
/function="putative regulator; Biosynthesis of cofactors, carriers: Molybdopterin"
/note="Residues 3 to 962 of 963 are 32.44 pct identical to residues 230 to 1209 of 1210 from Escherichia coli K-12 Strain MG1655: B2118"
/codon_start=1
/transl_table=11

```

/product-"putative regulator (fragment)"
/db_xref="GI:12516332"
/translation="WTWLRVLTDDTALSTVEKYDFPPLLYRDFRNFRAVLAMLANNG
VRGVSRIILLETEDHSDNPTLYLPERISSETENLYKWLKTNHPDAIOILLGVLGKKG
LEYLSKQSKHPAAIAAYATULAHEOKERKALVKLITATPELVCDVIMVWNAKAA
GLSLSRQSAVECEYATVMDPELLVSPMTKEKKNTPEVDFLPVLPVPSVSDVT
PEITKRLTYLTHFOQAQQATKQTLDTLPPIKKASMEKHLIPTPEQOILWLH
GPEKRESGEIKYEPAPQSAVDALLRFPALNAEFVHHNNAYKSNWLTALCYLP
GOQATSEINOIVKEDNYSSEGNLAIFGSAAIPAFMACLODRPRLCEPEFLGSEL
ALPMAQOQKKNYSYEDARNWLDYPRHAAAGLLPVALGKKGKORCAQALRLVNLN
QRETEAAGTQNDVLAALAFDSDPLEEYPAKIAPLPFGYFTLWRRPRTKSN
LPLSDAMRHGTLMSFRDITPAYAGLDIIRFETRESLAEEGWLIVFAWTEAGAPK
ENWATSGILGENDTARKLTPLIRAWPQESQHRKRAVSGLDVADIGSDVALMPLRTV
ARKIKFALQEHAREKINIVAEENGLTMAELEDRLAPDLGDDSSGLILDGPRKFTV
GDETLKPVCDANGKVLKDLKPKNQSDKQATQAVNLFKQLKDVRAIASQOYDRL
EQAMCQRRTAEQRLVEHPLVRLHTRRLRGVYNDENALITCFRVAEDSTSDA
QDELTLPAGNITGPHVLEISPEASAAFRQIADYELLPPFOOLBERGSGYHLADNER
HELTWQGRLOCAQIRIVGLRRGWRLEESGVYAMRKTTPHGDLELETFEFLIYE
TGYDQHPVSEYKITSPDDRYGQSSLTFSMLDDITASELINDIESLFD"
3923..5203
/gene="molR_C"
/notes="23285"
3923..5203
/gene="molR_C"
/function-"putative regulator; Biosynthesis of cofactors,
carriers: Molybdopterin"
/notes="Residues 100 to 402 of 426 are 43.03 pct identical
to residues 5 to 309 of 1210 from Escherichia coli K-12
Strain MG1655: B2118"
/codon_start=1
/transl_table=11
/product-"putative regulator (fragment)"
/db_xref="GI:12516333"
/translation="WKTYYQDEKSHKFWAVEQOQNELHINWKGVTGQSOVKSFAD
THAAKALAEKRVKGYEQVDANASHSDAKKVLVTEERKASINRQATNGLPW
LADNDPIILPNNIARHLSHRMPGEVPTKPEKQNLISRLANRTHYKYNLTDFY
SCSDWADIAQIAISLIKAQRTPEARTMAYTLEICFCQGMNDETSSLELDIVS
TFGIVAYETVFWQIDFDYKDEHLTFINADSKDSYRRNRNDFSLRLKRLHS
LAEERWONCKAKLVADPLDISLYOPLIAIILMPEIPEVAHEIVRLHVOADVPLEM
LKLVDPTLEILENYOYIDVENTYGASWSNTVLRQEGIAALRLTPYAEEDNCGDV
LKCINHPATINKKCKRFSFNCDSGISRSGKR"
5169..6908
/gene="molR_D"
/notes="23286"
5169..6908
/gene="molR_D"
/function-"putative regulator; Biosynthesis of cofactors,
carriers: Molybdopterin"
/notes="Residues 265 to 579 of 579 are 94.92 pct identical
to residues 19 to 333 of 333 from Escherichia coli K-12
Strain MG1655: B2117"
/codon_start=1
/transl_table=11
/product-"putative regulator (fragment)"
/db_xref="GI:12516334"
/translation="WALAEALVKDKKRLTLLVSLNANPLFVEKVLIGLSQAANL
LQOQOLLNRSNEDANSEILPMVLVSPWETONKNDNSPVGKYPKIPPLPYQOP
PLWTRPKLNSKSVLPDLSALQHLGLENLRFPOEEALYGLLQVKDADSLAEFAMD
FWTLTAPAGSKGSAWFTALGVGLGNDTDLPTLIRAWPQESQHRKRAVGLDILAAI
GSDIALQNGIAQKLFKPAQOERAKXKIAGTRESRELTVAELDRAPLDLDGDS
LLLDGFSQVTSFEDTLKPRVYDSSGRLKDLKPKNSDDSEQANDVAVNRYKLLKCD
ARTVAQOQVLESATKRRRWSPENFOLFVHEPLVRHLTRRLRWGVYSTENLLTCK
PRVADSVYSTADDDLTLPEDGISVPIHVLISPTDATAFGOLFADYELLPPRQL
DENSVALTEARNASALTWRAGKPCPSRVMLGANKIKCTPQDQGGYHGWIKPLGR
WSLMEIDEGFVAGSPAEALGAEQILSKLNLWEGXAESYSGWSNSTQEAQFSLVDAIT
ASELINDIEALFE"
6918..10547
/gene="yehi"
/notes="23287"
6918..10547

```

```

/gene="yehi"
/function-"putative regulator; Not classified"
/notes="Residues 1 to 1209 of 1209 are 94.04 pct identical
to residues 1 to 1210 of 1210 from Escherichia coli K-12
Strain MG1655: B2118"
/codon_start=1
/transl_table=11
/product-"putative regulator"
/db_xref="GI:12516335"
/translation="MDKELPWLADNAQLELYKKKTPLSHRNWPGEVSVITGSLIQ
TLGDELLQOQAKENITWYDKCSLEWQSAIQOAINLTGHEKPSIPALTMALICIAQ
NDSQQLLEIVQOEGLEVAVIVISQCIARRYESDSLVVTLOYQDEYGYGSGATY
NDFDLRLKHLISLAESCWQRCADKLIALPGIKRPFIALILPEKPEIAELASL
ESSRSLSHKSKWLKVATDNTAVKLERWGLDVFSDREASVMSQENRFGYACASLL
RQGLAAVPRLAMYAHKEDCGSLVQINHPVIRTLILLVADNKRSLQVAKYSNFP
HTLAALAEALALKEPPARPPIEDNKLPAQKAREYWRTLQIOTLMSQPOLAEE
VMSLSTQARAVLKSYLSALPKPVDGTNSNLPEILVSPWRSKKMTAPDLIARL
ELTPQVYQPGGERLAAETESAYFSTESLAQMEQKSGRVQLQELFGDDVWFLNY
ILPGDKDAARNSLIVQWYHYYQGRVEEILNGNSPEQAALAEQALRGHIELINWED
NYSRYRPEKSVNLYLLAQLPREMALTWELRINEKKHLFPAGEDYFLSLGLDALPCLL
LAFSHRPKETPELILNFGATELALPVARVRRFAQORDLARQWILQWPHETALIP
VFTKSDKSEAAALLALRLIYEHGHELLQTVANRQRTDVPFALEHLLKQGPWEIIPA
RIPKAPDFWHPQMSRPLITNNQPVTDALIEIGEMLFTQGGRFYSGLEQKTFQ
POTLAFAWDLFTAQOQAPAKDNMAFLALSLFGDESTARDLTQILAWPOGKSARA
VSGNLTLTMNNDMALIQLHHSQRAKSRPLRDNAEEFLQVVAENRGLSQEELADRLV
PTLGDDPOALSDFDFGQFTVRFDENLNPVIFDOQNVQKSVPRRLRADDOLKAPEA
LARLAKGLKDATQVSKNLLPRLEAALRTTRWSLADFLHSLFVNHPTRLVTTQRLNGV
YPANPRLLNAPRVAAEGGFCNAQDEPIDLADALIGIAHPLEMTAENRSFQAQFA
DYEIMPPFQLARRTLLTPDESTLRWEGKATVGQLMKRYKGWESYEDAFV
YDLGEYRLVLKFPNGFNHYNDKSKALMSFRSURYRDNKNSVTFFAELDFNLSEALSP
DVIFH"
10609..10926
/gene="yehk"
/notes="23288"
10609..10926
/gene="yehk"
/function-"orf; Unknown function"
/notes="Residues 1 to 105 of 105 are 99.04 pct identical to
residues 1 to 105 of 105 from GenPept 118 :
gi1405847|gb|AAA60479.1| (U00007) yehk [Escherichia coli]"
/codon_start=1
/transl_table=11
/protein_id="AAG57181.1"
/db_xref="GI:12516336"
/translation="MIVOKELVAIDYEVVPEDPFPFRLHKKSELTGSGVYRLER
FELRSTFHQRDREDADPLINDALIIRDECIDERKLGRGSEPTVIAFNELQNIFNQ
EIE"
10943..>11010
/notes="O-island #91; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
BASE COUNT 3013 a 2709 c 2653 g 2631 t 4 others
ORIGIN

Alignment Scores:
Pred. No.: 213 Length: 11010
Score: 85.50 Matches: 33
Percent Similarity: 36.91% Conservative: 22
Best Local Similarity: 22.15% Mismatches: 65
Query Match: 12.65% Indels: 29
DB: 1 Gaps: 3

US-10-048-196-2 (1-134) x AE005438 (1-11010)
Qy 6 TyrValThrAlaThrLeuMetThrAlaPhe---ThrLeuAlaSerCysAlaSerThrPro 24
Db 1716 TGGGTGACAAGGTAAGTACGGGTCACTTTTGGTGAATCTCTCGAGCTGACATCAGAAACA 1657
Qy 25 GluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaVal 44
Db 1656 GAAGGAACGGAAGCACTGGAAGATCAACACACAGCGGTATTTTCTTTCTTTCTTTGGTC 1597
Qy 45 LysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyTrpLysIleAlaAlaMetLys 64

```

```

Db 1596 ATCCAGGCGCGAACAACAGCAGCTCCGCAAC----- 1564
QY 65 LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLys 84
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1563 ATATCCACAGTGCATATTCACATCTTCAGGAGCTGCTGGGACGACATTCAGATAG 1504
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 85 AlaProGluAlaValSerLeuThrGluLeuThrAlaAlaIleAsnProAsnAlaArg 104
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1503 ATGCGTGGCGCTTTAGCATTAACCCAGGAATGAGCTGCATACACTCTGGCGTGGCG 1444
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 105 ThrGluValAlaGlnLysIleValArgHisSerLeuLysProCys----- 119
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1443 GTAATCAGTTTCACAGCGCTTACGCCACTCTTTATCTTCATGTATTCGCCAGAAAGTA 1384
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 120 -----MetLeuGluThrValAsn 125
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1383 GCATAAGCGCAATAGCGCAGCGGATGCTTTGTGCTGCTGCTTAAGTATTCACGAG 1324
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 126 AlaPheIleValProThrThrArg 134
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1323 TGCCTTTTCTTACCAATTACACCGAGG 1297
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

RESULT 39
AC123926/c
LOCUS
DEFINITION Mus musculus chromosome UNK clone RP24-326N11, WORKING DRAFT
ACCESSION AC123926
VERSION AC123926.1 GI:21326382
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 157377)
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 157377)
TITLE McPherson,J.D. and Waterston,R.H.
JOURNAL Direct Submission
AUTHORS Submitted (05-JUN-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0326N11
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 156063 bases at least Q40
Consensus quality: 156847 bases at least Q30
Consensus quality: 157390 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 158206; sum-of-contigs
Quality coverage: 13.54 in Q20 bases; agarose-fp
Quality coverage: 13.67 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

```

* as soon as it is available and the accession number will
* be preserved.
* 1 1149: contig of 1149 bp in length
* 1150 1249: gap of unknown length
* 1250 1436: contig of 3187 bp in length
* 4437 4536: gap of unknown length
* 4537 10728: contig of 6192 bp in length
* 10729 10828: gap of unknown length
* 10829 22669: contig of 11841 bp in length
* 22670 22769: gap of unknown length
* 22770 34710: contig of 11941 bp in length
* 34711 34810: gap of unknown length
* 34811 157377: contig of 122567 bp in length.

FEATURES
         source
         1..157377
         /organism="Mus musculus"
         /db_xref="taxon:10090"
         /chromosome="UNK"
         /clone="RP24-326N11"
         1..1149
         /note="assembly_name:Contig5"
         1250..4436
         /note="assembly_name:Contig12"
         4537..10728
         /note="assembly_name:Contig13"
         10829..22669
         /note="assembly_name:Contig14"
         22770..34710
         /note="assembly_name:Contig15"
         34811..157377
         /note="assembly_name:Contig16"
BASE COUNT 41968 a 34599 c 36427 g 43883 t 500 others
ORIGIN

Alignment Scores:
Pred. No.: 5,86e+03 Length: 157377
Score: 85.50 Matches: 25
Percent Similarity: 40.96% Conservative: 9
Best Local Similarity: 30.12% Mismatches: 42
Query Match: 12.65% Indels: 7
DB: 2 Gaps: 2

US-10-048-196-2 (1-134) x AC123926 (1-157377)

QY 23 ThrProGluSerAsnProLysAsnSerSerGluSerLysAlaLysIleSerGluThrAla 42
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 49987 AGCCGAGACAGAACCCCAATCCATCAGCAGCAAACTTAACAGTAAGTATGTGTGCAC 49928
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 43 AlaValLysGlnThrCysGlnThr-----GlnLeuThrGlyHisGlnTyr 57
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 49927 ACGTGCACACATGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 49868
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 58 TrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAla 77
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 49867 GGTACACGCTCTCTCTCCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 49808
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 78 CysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrAla 97
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 49807 TTTGTCTGTATAGCA-----CCTGTCGCCAGTGTGTTCTCTGCTGCTGCTGCTGCT 49754
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
QY 98 AlaIleAsn 100
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 49753 GCCATTAAT 49745
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

RESULT 40
LOCUS AC124403
DEFINITION Mus musculus chromosome UNK clone RP24-369I6, WORKING DRAFT
ACCESSION AC124403
VERSION AC124403.1 GI:21426522
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.

```

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	1 (bases 1 to 243098)
TITLE	McPherson,J.D. and Waterston,R.H.
JOURNAL	The sequence of Mus musculus clone
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 243098)
TITLE	McPherson,J.D. and Waterston,R.H.
JOURNAL	Direct Submission
REFERENCE	Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS	3 (bases 1 to 243098)
TITLE	McPherson,J.D. and Waterston,R.H.
JOURNAL	Direct Submission
COMMENT	Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

[illegible]

Search completed: May 5, 2003, 23:05:20
Job time : 2107 secs

[illegible]

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 5, 2003, 22:28:35 ; Search time 43 Seconds
(without alignments)
955.690 Million cell updates/sec

Title: us-10-048-196-2

Perfect score: 676

Sequence: 1 MKKILVVTATLTAFATLASC.....SLKPCMLETNNAFIVPTTTR 134

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10048196/runat_28042003_151440_5608/app_query.fasta_1.327
-DB=Issued Patents.NA -OFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10048196.ecgn_1_1_32_erunat_28042003_151440_5608 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents.NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	88.5	13.1	5857	4	US-08-961-527-28
c 2	85.5	12.6	3823	4	US-09-453-702B-132
c 3	77	11.4	2324	4	US-09-724-864-21
c 4	77	11.4	12588	2	US-08-387-942C-1
c 5	76.5	11.3	894	2	US-08-467-963C-28
c 6	76.5	11.3	894	2	US-08-838-189D-28
c 7	76.5	11.3	894	3	US-08-852-344D-28
c 8	76.5	11.3	894	3	US-08-344-639E-28
c 9	76.5	11.3	920	2	US-08-467-963C-7
c 10	76.5	11.3	920	2	US-08-838-189D-7
c 11	76.5	11.3	920	3	US-08-852-344D-7
c 12	76.5	11.3	920	3	US-08-344-639E-7

Sequence 7, Appli	76.5	11.3	920	4	US-08-467-969A-7	Sequence 7, Appli
Sequence 7, Appli	76.5	11.3	920	4	US-08-467-961A-7	Sequence 7, Appli
Sequence 7, Appli	76.5	11.3	920	4	US-08-001-554A-7	Sequence 152, App
Sequence 152, App	76.5	11.3	2179	4	US-09-370-838-152	Sequence 1, Appli
Sequence 1, Appli	75	11.1	4291	4	US-09-351-200-1	Sequence 98, Appli
Sequence 98, Appli	74	10.9	439	4	US-09-060-756-98	Sequence 1, Appli
Sequence 1, Appli	74	10.9	1712	2	US-08-632-598-1	Sequence 3, Appli
Sequence 3, Appli	74	10.9	1712	4	US-09-231-240-1	Sequence 28, Appli
Sequence 28, Appli	73.5	10.9	1080	4	US-09-125-642C-3	Sequence 31, Appli
Sequence 31, Appli	73	10.8	2574	4	US-09-255-829-28	Sequence 311, App
Sequence 311, App	73	10.8	6321	4	US-09-221-017B-311	Sequence 71, Appli
Sequence 71, Appli	73	10.8	6749	2	US-08-658-665-71	Sequence 121, App
Sequence 121, App	73	10.8	6749	4	US-09-085-273-71	Sequence 8, Appli
Sequence 8, Appli	72.5	10.7	5515	4	US-09-453-702B-121	Sequence 12, Appli
Sequence 12, Appli	72.5	10.7	5519	4	US-09-125-642C-8	Sequence 3, Appli
Sequence 3, Appli	72	10.7	33529	4	US-09-144-085-3	Sequence 1, Appli
Sequence 1, Appli	71.5	10.6	2490	1	US-08-171-382-1	Sequence 1, Appli
Sequence 1, Appli	71.5	10.6	2490	1	US-08-309-420-1	Sequence 1, Appli
Sequence 1, Appli	71.5	10.6	2490	1	US-08-309-419-1	Sequence 1, Appli
Sequence 1, Appli	71.5	10.6	2490	5	PCT-US95-11856-1	Sequence 1, Appli
Sequence 1, Appli	71.5	10.6	2490	5	PCT-US95-11878-1	Sequence 14, Appli
Sequence 14, Appli	70	10.4	794	3	US-09-010-809-14	Patent No. 5258502
Patent No. 5258502	70	10.4	2214	6	5258502-1	Sequence 1, Appli
Sequence 1, Appli	69.5	10.3	2793	1	US-08-209-747-1	Sequence 2, Appli
Sequence 2, Appli	69.5	10.3	2793	1	US-08-458-298-1	Sequence 31, Appli
Sequence 31, Appli	69.5	10.3	4565	3	US-08-776-265-2	Sequence 31, Appli
Sequence 31, Appli	69.5	10.3	6756	2	US-08-151-574-31	Sequence 31, Appli
Sequence 31, Appli	69.5	10.3	6756	2	US-08-419-448-31	Sequence 31, Appli
Sequence 31, Appli	69.5	10.3	6756	4	US-09-233-510-31	Sequence 3, Appli
Sequence 3, Appli	68.5	10.1	2702	3	US-08-508-761B-3	Sequence 263, App
Sequence 263, App	68.5	10.1	3744	4	US-08-961-527-263	Sequence 6, Appli
Sequence 6, Appli	68.5	10.1	4208	4	US-09-214-278-6	

ALIGNMENTS

RESULT 1

US-08-961-527-28/c
; Sequence 28, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:


```
; SEQ ID NO 21
; LENGTH: 2324
; TYPE: DNA
; ORGANISM: Mouse
US-09-724-864-21

Alignment Scores:
Pred. No.: 5.41 Length: 2324
Score: 77.00 Matches: 27
Percent Similarity: 37.93% Conservative: 17
Best Local Similarity: 23.28% Mismatches: 36
Query Match: 11.39% Indels: 36
DB: 4 Gaps: 6

US-10-048-196-2 (1-134) x US-09-724-864-21 (1-2324)
QY 12 MetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSer 31
DB 1934 GTGACTGCACCGCAATTCTGTTGTTCAAGTCTACCA-----CCAAGGACCTTC 1984
QY 32 SerAlaAsnLeuThrSerLeu-----IleLysHisAla----- 43
DB 1985 TGTTCAGGATGACACAAATGTTTCGTTAACTTCCAGAGGTACACACCTGAAAAAT 2044
QY 44 -----ValLysGlnThrCysGlnThrGlnLeuThr-----GlyHisGln 56
DB 2045 ACTTAGGAGCGAGTACATGCAATCTGCGTAACTGAGGAAGTCTCAACCTCAGCAG 2104
QY 57 TyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThr 76
DB 2105 TCTTGAGGCTGCACTT-----TCCACAACATTAAATCCAGAGGTG 2149
QY 77 AlaCysGlyCysValAlaAspLys----- 84
DB 2150 GGTGCCACTGGTGGAGACAGAGCTCCCTCCGTCGCCCATGGCTTCCTTGGTCT 2209
QY 85 ---AlaProGluAlaValSerLeuThrGluLeuThrAlaAlaIle 99
DB 2210 TCATGCCCTGAGGGTGGGCTAACTGGTGTAGTCTTGGCTGTG 2257

RESULT 4
US-08-387-942C-1
; Sequence 1, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: ERTESVAG, HELGA
; APPLICANT: VALLA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,942C
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1809-106P
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12588 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Azotobacter vinelandii
; STRAIN: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 290..1951
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2227..6438
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6702..9695
; NAME/KEY: CDS
; LOCATION: 9973..12588
; US-08-387-942C-1

Alignment Scores:
Pred. No.: 73.5 Length: 12588
Score: 77.00 Matches: 30
Percent Similarity: 36.21% Conservative: 12
Best Local Similarity: 25.86% Mismatches: 44
Query Match: 11.39% Indels: 30
DB: 2 Gaps: 4

US-10-048-196-2 (1-134) x US-08-387-942C-1 (1-12588)
QY 8 ThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsn 27
DB 11525 ACGGCAACTTCGTGGCCAGTTCAACGATGGCAACTGTTGTTCGACGCCCTCCGGTCA 11584
QY 28 ProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGlnThr 47
DB 11585 CCG-----GTACCGAGGGCAAGACA-----ACC 11608
QY 48 CysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLysLeuSer 67
DB 11609 TGTCCGGCACCGATCGCGGGAAACCCCTCTGGCTACGCGCGCAACGACACCTCAACG 11668
QY 68 GluSerLysAlaLysIleSerGluThrAla----- 77
DB 11669 GCGGGCGCGGCAACGACATCCTGTCGGCGCGCGCGGCGGCGGCGGCGGCGG 11728
QY 78 -----CysGly-----CysValAlaAspLysAlaProGluAlaVal 89
DB 11729 CCGGGCGCGGACGTGTTCCGCTTCGAGGCGCTGTCGACAGCCGCGCACTACACGCCG 11788
QY 90 SerLeuThrGluLeuThrThrAlaIleAsnProAsnAlaArgThr 105
DB 11789 GCGACAACCGAGCGGATTCATCATCGACTTCGCCGTGGCGGAAGACA 11836

RESULT 5
US-08-467-963C-28
; Sequence 28, Application US/08467963C
; Patent No. 5968776
; GENERAL INFORMATION:
; APPLICANT: KLEIN, Michel H
; APPLICANT: DU, Run-Pan
; APPLICANT: EWASYSHYN, Mary E
; TITLE OF INVENTION: MULTIMERIC HYBRID GENE ENCODING A
; TITLE OF INVENTION: CHIMERIC PROTEIN WHICH CONFERS PROTECTION AGAINST
; TITLE OF INVENTION: PARAINFLUENZA VIRUS AND RESPIRATORY SYNCYTIAL VIRUS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,963C
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/838,189
FILING DATE: 16-APR-1997
APPLICATION NUMBER: US 08/001,554
FILING DATE: 06-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9200117.1
FILING DATE: 06-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-474 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-467-963C-28

Alignment Scores:
Pred. No.: 1.44 Length: 894
Score: 76.50 Matches: 34
Percent Similarity: 40.44% Conservative: 21
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 11.32% Indels: 28
DB: 2 Gaps: 5

US-10-048-196-2 (1-134) x US-08-467-963C-28 (1-894)

QY 1 MetMetLysIleLeuThrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
Db 133 ATTCTGGCAATGATATCTCAACTTCACTTATTAATACAGCCCATCATATTCATAGCCTCG 192
QY 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
Db 193 GCA-----AACCACAAAGTCACACTAACAACCTGCAATCATA 228
QY 41 LysHisAla-----ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
Db 229 CAAGATGCAACAAAGCCAGATCAAGAACACAAACCAACATACCTCAGGATCTCTCAG 288
QY 58 TrpLysIleAlaAlaMetLysLeuSer-----SerGluSerLysAlaLysIleSer 74
Db 289 CTGTGGAATGAGTCTTCCATCTGCTGAATTTACATCAACAAACCAACCATAGCTAGCT 348
QY 75 GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeu 94
Db 349 TCACACACACAGGAGTCAAGTCAACCTGCAACCCCAACACAGCTCAAGACTATAAACA 408
QY 95 ThrThrAlaAlaIle-AsnPro-----AsnAlaArgThrGluValAl 108
Db 409 ACAACAACCAACACACAAACCCCAAGCCCACTACAAACACAGCCCAACA----- 460
QY 108 aGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThr 123

Db 461 -----AACCACCAAAACCAATAATGATTTCACT 493

RESULT 6

US-08-838-189D-28
Sequence 28, Application US/08838189D
Patent No. 5998169
GENERAL INFORMATION:

APPLICANT: KLEIN, Michel H
APPLICANT: DU, Run-Pan
APPLICANT: ENASYSHYN, Mary E
TITLE OF INVENTION: CHIMERIC HYBRID GENE ENCODING A
TITLE OF INVENTION: CHIMERIC PROTEIN WHICH CONFERS PROTECTION AGAINST
TITLE OF INVENTION: PARAINFLUENZA VIRUS AND RESPIRATORY SYNCYTIAL VIRUS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/838,189D
FILING DATE: 16-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/001,554
FILING DATE: 06-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9200117.1
FILING DATE: 06-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-687 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-838-189D-28

Alignment Scores:
Pred. No.: 1.44 Length: 894
Score: 76.50 Matches: 34
Percent Similarity: 40.44% Conservative: 21
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 11.32% Indels: 28
DB: 2 Gaps: 5

US-10-048-196-2 (1-134) x US-08-838-189D-28 (1-894)

QY 1 MetMetLysIleLeuThrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
Db 133 ATTCTGGCAATGATATCTCAACTTCACTTATTAATACAGCCCATCATATTCATAGCCTCG 192
QY 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
Db 193 GCA-----AACCACAAAGTCACACTAACAACCTGCAATCATA 228
QY 41 LysHisAla-----ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
Db 229 CAAGATGCAACAAAGCCAGATCAAGAACACAAACCAACATACCTCAGGATCTCTCAG 288

```
QY 58 TrpLysIleAlaAlaMetLysLeuSer-----SerGluSerLysAlaLysIleSer 74
|||||
Db 289 CTTGGAATCAGCTTCTCCAAATCTGCTGAATTAATACATAACCAACCACTACTAGCT 348
QY 75 GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeu 94
|||||
Db 349 TCAACAACACCCAGGAGTCAAGTCAAACTGCAACCCCAACCAAGTCAAGACTAAACACACA 408
QY 95 ThrThrAlaAlaIle-AsnPro-----AsnAlaArgThrGluValAl 108
|||||
Db 409 ACAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 460
QY 108 aGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThr 123
|||||
Db 461 -----AACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 493

RESULT 7
US-08-852-344D-28
; Sequence 28, Application US/08852344D
; Patent No. 6017539
; GENERAL INFORMATION:
; APPLICANT: KLEIN, Michel H
; APPLICANT: DU, Run-Pan
; APPLICANT: EWASYSHYN, Mary E
; TITLE OF INVENTION: CHIMERIC PROTEIN WHICH CONFERS PROTECTION
; TITLE OF INVENTION: AGAINST PARAINFLUENZA VIRUS AND RESPIRATORY
; TITLE OF INVENTION: SYNCYTIAL VIRUS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,344D
; FILING DATE: 07-MAY-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,639
; FILING DATE: 14-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9200117.1
; FILING DATE: 06-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-688 MIS:jb
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-852-344D-28

Alignment Scores:
Pred. No.: 1.44
Score: 76.50
Percent Similarity: 40.44%
Best Local Similarity: 25.00%
Query Match: 11.32%
DB: 3
Length: 894
Matches: 34
Conservative: 21
Mismatch: 53
Indels: 28
Gaps: 5
```

```
US-10-048-196-2 (1-134) x US-08-852-344D-28 (1-894)
QY 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
|||||
Db 133 ATTCTGGCANTGTAATCTCAACTTCACTTATAATTACAGCCATCATATTATCATAGCTCG 192
QY 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
|||||
Db 193 GCA-----AACCACAAAGTCACACTAACCACTGCAATCAT 228
QY 41 LysHisAla-----ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
|||||
Db 229 CAAGATGCAACCAAGCCAGATCAAGACACACCAACCCCAACATACCTCTACTCAGGATCTCAG 288
QY 58 TrpLysIleAlaAlaMetLysLeuSer-----SerGluSerLysAlaLysIleSer 74
|||||
Db 289 CTTGGAATCAGCTTCTCCAAATCTGCTGAATTAATACATAACCAACCACTACTAGCT 348
QY 75 GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeu 94
|||||
Db 349 TCAACAACACCCAGGAGTCAAGTCAAACTGCAACCCCAACCAAGTCAAGACTAAACACACA 408
QY 95 ThrThrAlaAlaIle-AsnPro-----AsnAlaArgThrGluValAl 108
|||||
Db 409 ACAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 460
QY 108 aGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThr 123
|||||
Db 461 -----AACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 493

RESULT 8
US-08-344-639E-28
; Sequence 28, Application US/08344639E
; Patent No. 6033668
; GENERAL INFORMATION:
; APPLICANT: Klein, Michel H
; APPLICANT: Du, Run-Pan
; APPLICANT: Ewasysyn, Mary E
; TITLE OF INVENTION: CHIMERIC PROTEIN WHICH CONFERS
; TITLE OF INVENTION: PROTECTION AGAINST PARAINFLUENZA VIRUS
; TITLE OF INVENTION: AND RESPIRATORY SYNCYTIAL VIRUS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6th Floor
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/344,639E
; FILING DATE: 14-NOV-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/001,554
; FILING DATE: 06-JAN-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9200117.1
; FILING DATE: 06-JAN-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-391 MIS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
```


;; TITLE OF INVENTION: CHIMERIC PROTEIN WHICH CONFERS PROTECTION AGAINST
;; TITLE OF INVENTION: PARAINFLUENZA VIRUS AND RESPIRATORY SYNCYTIAL VIRUS
;; NUMBER OF SEQUENCES: 38
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sim & McBurney
;; STREET: 6th Floor, 330 University Avenue
;; CITY: Toronto
;; STATE: Ontario
;; COUNTRY: Canada
;; ZIP: M5G 1R7
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/838,189D
;; FILING DATE: 16-APR-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/001,554
;; FILING DATE: 06-JAN-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: GB 9200117.1
;; FILING DATE: 06-JAN-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: STEWART, Michael I
;; REGISTRATION NUMBER: 24,973
;; REFERENCE/DOCKET NUMBER: 1038-687 MIS:jb
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (416) 595-1155
;; TELEFAX: (416) 595-1163
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 920 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-838-189D-7

Alignment Scores:
Pred. No.: 1.5 Length: 920
Score: 76.50 Matches: 34
Percent Similarity: 40.44% Conservative: 21
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 11.32% Indels: 28
DB: 2 Gaps: 5

US-10-048-196-2 (1-134) x US-08-838-189D-7 (1-920)

QY 1 MetMetLysIleValArgHisSerLeuLysProCysMetLeuGluThr 123
Db 140 ATCTGGCAATGATATCTCAACTTCACTATATAATTACAGCCATCATATTCATAGCCTCG 199
QY 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
Db 200 GCA-----AACCACAAAGTCACACTAACACCAACCACTCAGGATCCTCAG 295
QY 41 LysHisAla-----ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
Db 236 CAAGATGCAACAGCCAGATCAAGAACACACACCAACCACTCACTCAGGATCCTCAG 295
QY 58 TrpLysIleAlaAlaMetLysLeuSer-----SerGluSerLysAlaLysIleSer 74
Db 296 CTGTGAATCAGCTCTCTCCCAATCTGTGTAATTTACATCAACCAACCACTAGCT 355
QY 75 GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeu 94
Db 356 TCAACACACAGGATCAGTCAAGTCAACACCTGCAACCCCAACCACTCAAGACTTAAACAC 415
QY 95 ThrThrAlaAlaIleAsnPro-----AsnAlaArgThrGluValAl 108
Db 416 ACAACACCCCAACACACCAACCCAGCCCACTACAAACCAACGCAACCAACA----- 467

QY 108 aGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThr 123
Db 468 -----AACCACCAAAACCAACCAATGATTTCACT 500
RESULT 11
US-08-852-344D-7
;; Sequence 7, Application US/08852344D
;; Patent No. 6017539
;; GENERAL INFORMATION:
;; APPLICANT: KLEIN, Michael H
;; APPLICANT: DU, Run-Pan
;; APPLICANT: EWASYSHYN, Mary E
;; TITLE OF INVENTION: CHIMERIC PROTEIN WHICH CONFERS PROTECTION
;; TITLE OF INVENTION: AGAINST PARAINFLUENZA VIRUS AND RESPIRATORY
;; TITLE OF INVENTION: SYNCYTIAL VIRUS
;; NUMBER OF SEQUENCES: 37
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sim & McBurney
;; STREET: 6th Floor, 330 University Avenue
;; CITY: Toronto
;; STATE: Ontario
;; COUNTRY: Canada
;; ZIP: M5G 1R7
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/852,344D
;; FILING DATE: 07-MAY-1997
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/344,639
;; FILING DATE: 14-NOV-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: GB 9200117.1
;; FILING DATE: 06-JAN-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: STEWART, Michael I
;; REGISTRATION NUMBER: 24,973
;; REFERENCE/DOCKET NUMBER: 1038-688 MIS:jb
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (416) 595-1155
;; TELEFAX: (416) 595-1163
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 920 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-852-344D-7

Alignment Scores:
Pred. No.: 1.5 Length: 920
Score: 76.50 Matches: 34
Percent Similarity: 40.44% Conservative: 21
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 11.32% Indels: 28
DB: 3 Gaps: 5

US-10-048-196-2 (1-134) x US-08-852-344D-7 (1-920)

QY 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
Db 140 ATCTGGCAATGATATCTCAACTTCACTATATAATTACAGCCATCATATTCATAGCCTCG 199
QY 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
Db 200 GCA-----AACCACAAAGTCACACTAACACCAACCACTCAGGATCCTCAG 295
QY 41 LysHisAla-----ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57

Db 236 CAAGATCAACACAGAGATCAAGAACACACACCCACATACCTCACTAGGATCCTCAG 295
Qy 58 TrpLysileAlaAlaMetLysLeuSer-----SerGluSerLysAlaLysIleSer 74
Db 296 CTTGGATCAGCTCTCCATCTGCTGAATTAATACATCAACACCCACCATCTAGCT 355
Qy 75 GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLysLeuThrGluLeu 94
Db 356 TCAACACACAGAGATCAAGTCAAACTGCAACCCACACAGCTCAAGACTAAAAACACA 415
Qy 95 ThrThrAlaAlaIle-AsnPro-----AsnAlaArgThrGluValAl 108
Db 416 ACAACACCCCAACACACAGCCAGCAAGCCCACTACAAACCAACCCCAACACA----- 467
Qy 108 aGlnLysIleValArgHisSerLysProCysMetLysLeuThr 123
Db 468 -----AACCCACCAACAAACCCCAATTAATGATTTTCACT 500

RESULT 12

US-08-344-639E-7
; Sequence 7, Application US/08344639E
; Patent No. 6033668
; GENERAL INFORMATION:
; APPLICANT: Klein, Michel H
; APPLICANT: Du, Run-Pan
; APPLICANT: Ewasysyn, Mary E
; TITLE OF INVENTION: CHIMERIC PROTEIN WHICH CONFERS
; TITLE OF INVENTION: PROTECTION AGAINST PARAINFLUENZA VIRUS
; TITLE OF INVENTION: AND RESPIRATORY SYNCYTIAL VIRUS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6th Floor
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/344,639E
; FILING DATE: 14-NOV-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/001,554
; FILING DATE: 06-JAN-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9200117.1
; FILING DATE: 06-JAN-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-391 MIS
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; TELETYPE: 065-24567 SIMBAS
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-344-639E-7

Alignment Scores:

Pred. No.: 1.5 Length: 920
Score: 76.50 Matches: 34
Percent Similarity: 40.44% Conservative: 21
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 11.32% Indels: 28
DB: 3 Gaps: 5
US-10-048-196-2 (1-134) x US-08-344-639E-7 (1-920)
Qy 1 MetMetLysIleLeuTyValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
Db 140 ATTCTGCAATGATAATCTCAACTTCTATATAATTACAGCCATCATATTACATAGCCTCG 199
Qy 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
Db 200 GCA-----AACCAACAAAGTCACACTAACCACTGCAATCATATA, 235
Qy 41 LysHisAla-----ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
Db 236 CAAGATGCAACACAGCCAGATCAAGAACACACACCCCACTACCTCACTCAGGATCCTCAG 295
Qy 58 TrpLysIleAlaAlaMetLysLeuSer-----SerGluSerLysAlaLysIleSer 74
Db 296 CTTGGATCAGCTCTCCATCTGCTGAATTAATACATCAACACCCACCATCTAGCT 355
Qy 75 GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLysLeuThrGluLeu 94
Db 356 TCAACACACAGAGATCAAGTCAAACTGCAACCCACACAGCTCAAGACTAAAAACACA 415
Qy 95 ThrThrAlaAlaIle-AsnPro-----AsnAlaArgThrGluValAl 108
Db 416 ACAACACCCCAACACACAGCCAGCAAGCCCACTACAAACCAACCCCAACACA----- 467
Qy 108 aGlnLysIleValArgHisSerLysProCysMetLysLeuThr 123
Db 468 -----AACCCACCAACAAACCCCAATTAATGATTTTCACT 500

RESULT 13

US-08-467-969A-7
; Sequence 7, Application US/08467969A
; Patent No. 6168786
; GENERAL INFORMATION:
; APPLICANT: Klein, Michel H
; APPLICANT: Du, Run-Pan
; APPLICANT: Ewasysyn, Mary E
; TITLE OF INVENTION: Chimeric Immunogens
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6th Floor
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,969A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/001,554
; FILING DATE: 06-JAN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9200117.1
; FILING DATE: 06-JAN-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I


```

; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-475 MIS:bh
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; TELEX: 065-24567 SIMBAS
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-467-969A-7

```

```

Alignment Scores:
Pred. No.: 1.5 Length: 920
Score: 76.50 Matches: 34
Percent Similarity: 40.44% Conservative: 21
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 11.32% Indels: 28
DB: 4 Gaps: 5

```

US-10-048-196-2 (1-134) x US-08-467-969A-7 (1-920)

```

Qy 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
Db 140 ATTCGGAATGATAATCTCAACTTCACTTATAATTACAGCCATCATTTTCATAGCCTCG 199
Qy 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
Db 200 GCA-----AACCACAAAGTCACACTAACAACTGCAATCATA 235
Qy 41 LysHisAla-----ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
Db 236 CAAGATGCAACAGCCAGAGTCAAGTCAAACTGCAACCCACACAGTCAAGACTAAACACACA 295
Qy 58 TrpLysIleAlaAlaMetLysLeuSer-----SerGluSerLysAlaLysIleSer 74
Db 296 CTGGAATGAGTCTCTCAATCTCTGGAATTACATCAACAAACCCACCACTACTAGCT 355
Qy 75 GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeu 94
Db 356 TCACACACACAGGAGTCAAGTCAAACTGCAACCCACACAGTCAAGACTAAACACACA 415
Qy 95 ThrThrAlaLalle-AsnPro-----AsnAlaArgThrGluValAl 108
Db 416 ACAACAACCCCAACACACACAGCCAGCAGCCACTACAAACACGCCAAACACA----- 467
Qy 108 aGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThr 123
Db 468 -----AACCACCAACAAACCCCAATGATTTTCACT 500

```

RESULT 14

US-08-467-961A-7

```

; Sequence 7, Application US/08467961A
; Patent No. 6171783
; GENERAL INFORMATION:
; APPLICANT: Klein, Michel H
; APPLICANT: Du, Run-Pan
; APPLICANT: Ewasishyn, Mary E
; TITLE OF INVENTION: Chimeric Immunogens
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 330 University Avenue, 6TH Floor
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,961A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/001,554
; FILING DATE: 06-JAN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9200117.1
; FILING DATE: 06-JAN-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-475 MIS:bh
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; TELEX: 065-24567 SIMBAS
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-467-961A-7

```

```

Alignment Scores:
Pred. No.: 1.5 Length: 920
Score: 76.50 Matches: 34
Percent Similarity: 40.44% Conservative: 21
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 11.32% Indels: 28
DB: 4 Gaps: 5

```

US-10-048-196-2 (1-134) x US-08-467-961A-7 (1-920)

```

Qy 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
Db 140 ATTCGGAATGATAATCTCAACTTCACTTATAATTACAGCCATCATTTTCATAGCCTCG 199
Qy 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
Db 200 GCA-----AACCACAAAGTCACACTAACAACTGCAATCATA 235
Qy 41 LysHisAla-----ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
Db 236 CAAGATGCAACAGCCAGAGTCAAGTCAAACTGCAACCCACACAGTCAAGACTAAACACACA 295
Qy 58 TrpLysIleAlaAlaMetLysLeuSer-----SerGluSerLysAlaLysIleSer 74
Db 296 CTGGAATGAGTCTCTCAATCTCTGGAATTACATCAACAAACCCACCACTACTAGCT 355
Qy 75 GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeu 94
Db 356 TCACACACACAGGAGTCAAGTCAAACTGCAACCCACACAGTCAAGACTAAACACACA 415
Qy 95 ThrThrAlaLalle-AsnPro-----AsnAlaArgThrGluValAl 108
Db 416 ACAACAACCCCAACACACAGCCAGCAGCCACTACAAACACGCCAAACACA----- 467
Qy 108 aGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThr 123
Db 468 -----AACCACCAACAAACCCCAATGATTTTCACT 500

```

RESULT 15

US-08-001-554A-7

```

; Sequence 7, Application US/08001554A
; Patent No. 6225091

```



```
; Patent No. 6320033
; GENERAL INFORMATION:
; APPLICANT: BOURBONNAIS, Yves
; APPLICANT: LAMARRE, Claude
; APPLICANT: DESLAURIERS, No. 63200331a
; TITLE OF INVENTION: CANDIDA ALBICANS GENE (CSA1) ENCODING A
; FILE REFERENCE: 6013-71-US- CC/
; CURRENT APPLICATION NUMBER: US/09/351.200
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: CA2,237,134
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4291
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: gene
; LOCATION: (0)...(0)
; US-09-351-200-1

Alignment Scores:
Pred. No.: 25.2 Length: 4291
Score: 75.00 Matches: 36
Percent Similarity: 41.50% Conservative: 25
Best Local Similarity: 24.49% Mismatches: 54
Query Match: 11.09% Indels: 32
DB: 4 Gaps: 6

US-10-048-196-2 (1-134) x US-09-351-200-1 (1-4291)
Qy 8 ThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsn 27
Db 1625 ACTTCATCTGAAACTTCTCCCAACAACATTTCAAGTATCATCTTCAGCAGACACTCC 1684
Qy 28 ProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLeuLysHisAlaValLys----- 45
Db 1685 GCTACAAGCAGCTCCTCAACACCATCTACTTTATTAGAACTGCTTCCATTAAATGGTTT 1744
Qy 46 -----GlnThrCysGlnThrGlnLeuThr 53
Db 1745 GCTGATAAACTTTAGCACCATTACCAGAAATGTGTAACCATGATGATGTTCCAAATACT 1804
Qy 54 Gly-----HisGlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLys 70
Db 1805 GGCATAACACCATGTCATCTACTGGGATCGCGTGTGTTATGTCATGCCACAATTTGCA 1864
Qy 71 AlaLysIleSerGluThrAlaCysGlyCysValAlaAsp-----LysAlaProGluAla 88
Db 1865 GGTGCTATTGGT-----TCATGTGTGTCGCGATGTTGTAAGGTCAGATATT 1912
Qy 89 ValSerLeuThrGluLeuThrThrAlaAlaIleAsn-----ProAsnAlaAargThr 105
Db 1913 GTTCTGTCCACGAGTGGGTACTTCTGTTGTCGTGCGGTGTTAATGCACCTTAT 1972
Qy 106 GluValAlaGlnLysIleValAargHisSerLeuLysProCysMetLeuGluThrValAsn 125
Db 1973 TGGATGCTTCCAGCAGTGTAAAGTAGCTTA-----AGTGTGCT 2014
Qy 126 AlaPheIleValProThrThr 132
Db 2015 GCTACTGCGAGTACCAACCTCC 2035

RESULT 18
; Sequence 98, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen

; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 98
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; US-09-060-756-98

Alignment Scores:
Pred. No.: 1 Length: 439
Score: 74.00 Matches: 24
Percent Similarity: 48.24% Conservative: 17
Best Local Similarity: 28.24% Mismatches: 38
Query Match: 10.95% Indels: 6
DB: 4 Gaps: 2

US-10-048-196-2 (1-134) x US-09-060-756-98 (1-439)
Qy 18 AlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThr 37
Db 200 GCTCGCGCGGCTTCTCCGCGCATGGTAACCCCGAGTTTCGCCCCAGGCTCTTAGCTAGC 259
Qy 38 SerLeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGln--- 56
Db 260 AGTGGCAGCGGTGTTCATCTGCATCGACCTGTGTGATCATCTCTCAAGGATGAAGCG 319
Qy 57 TyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThr 76
Db 320 TACTGGCGGACTGCGCTTCTCGCGCGCCGCAATCACCAGGATCTCAGAAAGC 379
Qy 77 AlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThr 96
Db 380 AGCTCGCATCTACTCTTCAG-----GCCACCTCGCCGAAAGCTCGACA 424
Qy 97 AlaAlaIleAsnPro 101
Db 425 TGGTCAATCCGCGCC 439

RESULT 19
; US-08-632-598-1/c
; Sequence 1, Application US/08632598
; Patent No. 5886164
; GENERAL INFORMATION:
; APPLICANT: BIRD, COLIN R
; APPLICANT: FLETCHER, JONATHAN D
; TITLE OF INVENTION: RIPENING-RELATED GENES FROM BANANA
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABY AND CUSHMAN
; STREET: 1100 NEW YORK AVENUE N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/632,598
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
```


US-08-658-665-71

Alignment Scores:

Pred. No.: 91.8 Length: 6749
Score: 73.00 Matches: 26
Percent Similarity: 39.45% Conservative: 17
Best Local Similarity: 23.85% Mismatches: 45
Query Match: 10.80% Indels: 21
DB: 2 Gaps: 2

US-10-048-196-2 (1-134) x US-08-658-665-71 (1-6749)

QY 18 AlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThr 37
|||||
Db 1874 GCGTCTGGCCCTCGGTGGTTACGAGTCTCTCGTCGGAGTCGATCCGCTGCCAGACG 1815
QY 38 SerLeuLeuLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHis----- 55
|||||
Db 1814 TCGTGTGCGCCCTGGCGGCACCCCTCGTGTGCGCGTCCAGGTGTCGGTACTCAAGC 1755
QY 56 -----GlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAla 71
|||||
Db 1754 TTGCCCTGGATCGATACGATCGGTGGTG----- 1728
QY 72 LysIleSerGluThrAlaCysGly-CysValAlaAspLysAlaProGluAlaValSerLe 91
|||||
Db 1727 -----AAGGTGGGGTGTTCGCTGTACTGAGGCCGCGCTGCAGCAGCAGCTCGAT 1677
QY 91 uThrGluLeuThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysI 111
|||||
Db 1676 ATCGAAAAGAGACGCGACCGACGGGATCGTACTGACGACGATTCACCGGTCTCGGTAT 1617
QY 111 eValArgHisSerLeuLysProCys 119
|||||
Db 1616 CGCTTGCACCTCCAGGAAGATCTGC 1592

RESULT 25

US-09-085-273-71/c
; Sequence 71, Application US/09085273
; Patent No. 6267965

GENERAL INFORMATION:

; APPLICANT: Paolletti, Enzo
; APPLICANT: Pincus, Steven E.
; APPLICANT: Cox, William I.
; APPLICANT: Kauffman, Elizabeth K.
; TITLE OF INVENTION: RECOMBINANT POXVIRUS - CYTOMEGALOVIRUS,
; TITLE OF INVENTION: COMPOSITIONS AND USES
; NUMBER OF SEQUENCES: 176
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,273
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/471,014
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2720
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333

; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6749 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-085-273-71

Alignment Scores:

Pred. No.: 91.8 Length: 6749
Score: 73.00 Matches: 26
Percent Similarity: 39.45% Conservative: 17
Best Local Similarity: 23.85% Mismatches: 45
Query Match: 10.80% Indels: 21
DB: 2 Gaps: 2

US-10-048-196-2 (1-134) x US-09-085-273-71 (1-6749)

QY 18 AlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThr 37
|||||
Db 1874 GCGTCTGGCCCTCGGTGGTTACGAGTCTCTCGTCGGAGTCGATCCGCTGCCAGACG 1815
QY 38 SerLeuLeuLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHis----- 55
|||||
Db 1814 TCGTGTGCGCCCTGGCGGCACCCCTCGTGTGCGCGTCCAGGTGTCGGTACTCAAGC 1755
QY 56 -----GlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAla 71
|||||
Db 1754 TTGCCCTGGATCGATACGATCGGTGGTG----- 1728
QY 72 LysIleSerGluThrAlaCysGly-CysValAlaAspLysAlaProGluAlaValSerLe 91
|||||
Db 1727 -----AAGGTGGGGTGTTCGCTGTACTGAGGCCGCGCTGCAGCAGCAGCTCGAT 1677
QY 91 uThrGluLeuThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysI 111
|||||
Db 1676 ATCGAAAAGAGACGCGACCGACGGGATCGTACTGACGACGATTCACCGGTCTCGGTAT 1617
QY 111 eValArgHisSerLeuLysProCys 119
|||||
Db 1616 CGCTTGCACCTCCAGGAAGATCTGC 1592

RESULT 26

US-09-453-702B-121
; Sequence 121, Application US/09453702B
; Patent No. 6365723

GENERAL INFORMATION:

; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli 0157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/171.382
FILING DATE:

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: IRI430USA
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2490 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 205..2286
US-08-171-382-1

Alignment Scores:
Pred. No.: 30.7 Length: 2490
Score: 71.50 Matches: 30
Percent Similarity: 43.24% Conservative: 18
Best Local Similarity: 27.03% Mismatches: 41
Query Match: 10.58% Indels: 22
DB: 1 Gaps: 4

US-10-048-196-2 (1-134) x US-08-171-382-1 (1-2490)

Qy 25 GluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle----- 40
||| |||||: |||||: |||
Db 1838 GAAATAAACCTAGGATTTTCATCATCTACTTTTGGAACTGATAATATCTGTGTGCTCC 1779
Qy 41 -----LysHisAlaValLysGlnThrCys----- 48
||| |||||: |||||: |||
Db 1778 CTGGAAGAACTGGATATTTCGTCAGTGAAGCAGCTGCCTGCAACTTGTGCTCCAT 1719
Qy 49 GlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLysLeuSerGlu 68
::: |||||: |||||: |||
Db 1718 TCTTCTCAATTA-----TTTAAATAGATTCATGGAAGGAAAGATAGT 1671
Qy 69 SerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAla 88
||| |||||: |||||: |||
Db 1670 TCTGAGGCATCTATTTTA---TCATCTGTGATTGCTTAGCCACTTCTAGACCTAAGTA 1614
Qy 89 ValSerLeuThrGluLeuThrThrAlaAla--IleAsnProAsnAlaArgThrGluValAl 108
||| |||||: |||||: |||
Db 1613 GTGAGTGAATGAGAGACACAGCTTTTGGCAAAAGAGACATCAGTTCCGATCCAGGTATG 1554
Qy 108 aGlnLysIleValArgHisSerLeuLysPro 118
::: |||||: |||||: |||
Db 1553 TTCTGAATGCCACAAAGGAACTGAATCCC 1523

RESULT 31

US-08-309-420-1/c
; Sequence 1, Application US/08309420
; Patent No. 5591588
; GENERAL INFORMATION:

APPLICANT: Goldstein, Gideon
APPLICANT: Culler, Michael
TITLE OF INVENTION: Method for the Diagnosis of Depression
TITLE OF INVENTION: Based on Monitoring Blood Levels of Arginine Vasopressin
and/or Thymopoietin
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: 321 No. 5591588ristown Road, Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/309.420
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: IRI460USA
TELEPHONE: (215) 540-9207
TELEFAX: (215) 540-5818

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2490 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 205..2286
US-08-309-420-1

Alignment Scores:
Pred. No.: 30.7 Length: 2490
Score: 71.50 Matches: 30
Percent Similarity: 43.24% Conservative: 18
Best Local Similarity: 27.03% Mismatches: 41
Query Match: 10.58% Indels: 22
DB: 1 Gaps: 4

US-10-048-196-2 (1-134) x US-08-309-420-1 (1-2490)

Qy 25 GluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle----- 40
||| |||||: |||||: |||
Db 1838 GAAATAAACCTAGGATTTTCATCATCTACTTTTGGAACTGATAATATCTGTGTGCTCC 1779
Qy 41 -----LysHisAlaValLysGlnThrCys----- 48
||| |||||: |||||: |||
Db 1778 CTGGAAGAACTGGATATTTCGTCAGTGAAGCAGCTGCCTGCAACTTGTGCTCCAT 1719
Qy 49 GlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLysLeuSerGlu 68
::: |||||: |||||: |||
Db 1718 TCTTCTCAATTA-----TTTAAATAGATTCATGGAAGGAAAGATAGT 1671
Qy 69 SerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAla 88
||| |||||: |||||: |||
Db 1670 TCTGAGGCATCTATTTTA---TCATCTGTGATTGCTTAGCCACTTCTAGACCTAAGTA 1614
Qy 89 ValSerLeuThrGluLeuThrThrAlaAla--IleAsnProAsnAlaArgThrGluValAl 108
||| |||||: |||||: |||
Db 1613 GTGAGTGAATGAGAGACACAGCTTTTGGCAAAAGAGACATCAGTTCCGATCCAGGTATG 1554
Qy 108 aGlnLysIleValArgHisSerLeuLysPro 118
::: |||||: |||||: |||

```
Db 1553 TTCTGAATGCCACAAGAACCTGAATCCC 1523
RESULT 32
US-08-309-419-1/c
; Sequence 1, Application US/08309419
; Patent No. 5593842
; GENERAL INFORMATION:
; APPLICANT: Goldstein, Gideon
; APPLICANT: Culler, Michael
; TITLE OF INVENTION: Method of Measuring Thymopoietin
; TITLE OF INVENTION: Proteins in Plasma and Serum
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: 321 No. 5593842ristown Road, Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,419
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: IRI45BUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9206
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2490 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 205..2286
; US-08-309-419-1
Alignment Scores:
Pred. No.: 30.7 Length: 2490
Score: 71.50 Matches: 30
Percent Similarity: 43.24% Conservative: 18
Best Local Similarity: 27.03% Mismatches: 41
Query Match: 10.58% Indels: 22
DB: 1 Gaps: 4
US-10-048-196-2 (1-134) x US-08-309-419-1 (1-2490)
Qy 25 GluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeu----- 40
||| |||||::: |||||::: |||
Db 1838 GAAATAAACCCCTAGGATTTCATCTACTTTTGGAACTGATAATATCTGTGTTGCCCTCC 1779
Qy 41 -----LysHisAlaValLysGlnThrCys----- 48
||| |||||::: |||||::: |||
Db 1778 CTGGAAGAACTGGATATTTTCATGCGCCAGTGAAGCGACTGCTCAACTTGTGCTGCCAT 1719
Qy 49 GlnThrGlnLeuThrGlyHisGlnTrpLysIleAlaAlaMetLysLeuSerSerGlu 68
::: |||||::: |||||::: |||
Db 1718 TCTTCTCAATTA-----TTTAAATAGATTTCATGGAAGGGAAGAAAGATAGT 1671
69 SerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAla 88
|||::: |||||::: |||||::: |||
Db 1670 TCTGAGGCATCTATTTTA---TCATGCTGTGATTGCTTAGCCACTTCTAGACCTTAAGGTA 1614
Qy 89 ValSerLeuThrGluLeuThrThrAlaAla-IleAsnProAsnAlaArgThrGluValAl 108
||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1613 GTGAGTGAATGAGAGACAAACAGTTTGGCAAAAGAGACATCAGTTCCGATCCAGGTATG 1554
Qy 108 aGlnLysIleValAlaArgHisSerLeuLyspro 118
::: |||||::: |||||::: |||
Db 1553 TTCTGAATGCCACAAGAACCTGAATCCC 1523
RESULT 33
PCT-US95-11856-1/c
; Sequence 1, Application PC/TUS9511856
; GENERAL INFORMATION:
; APPLICANT: Immunobiology Research, Institute Inc.
; TITLE OF INVENTION: Method of Measuring
; TITLE OF INVENTION: Thymopoietin Proteins in Plasma and Serum
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: 321 Norristown Road, Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11856
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,419
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: IRI45BPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9206
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2490 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 205..2286
; PCT-US95-11856-1
Alignment Scores:
Pred. No.: 30.7 Length: 2490
Score: 71.50 Matches: 30
Percent Similarity: 43.24% Conservative: 18
Best Local Similarity: 27.03% Mismatches: 41
Query Match: 10.58% Indels: 22
DB: 5 Gaps: 4
US-10-048-196-2 (1-134) x PCT-US95-11856-1 (1-2490)
Qy 25 GluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeu----- 40
||| |||||::: |||||::: |||
Db 1838 GAAATAAACCCCTAGGATTTCATCTACTTTTGGAACTGATAATATCTGTGTTGCCCTCC 1779
Qy 41 -----LysHisAlaValLysGlnThrCys----- 48
||| |||||::: |||||::: |||
Db 1778 CTGGAAGAACTGGATATTTTCATGCGCCAGTGAAGCGACTGCTCAACTTGTGCTGCCAT 1719
Qy 49 GlnThrGlnLeuThrGlyHisGlnTrpLysIleAlaAlaMetLysLeuSerSerGlu 68
```

Db 1718 TCTTCTCAATTA---TTTAAATAGATTTCATGGAAGGAAAGATAGT 1671
QY 69 SerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAla 88
Db 1670 TCTGAGGATCTATTTTA---TCATGCTGTGATTGCTTAGCCACTTCTAGACCTTAAGGTA 1614
QY 89 ValSerLeuThrGluLeuThrThrAlaAla-IleAsnProAsnAlaArgThrGluValAla 108
Db 1613 GTGAGTGAATGAGACACAGCTTTTGGCAAAAGAGACATCAGTTCCGGATCCAGGTATG 1554
QY 108 aglnLysIleValArgHisSerLeuLysPro 118
Db 1553 TTCTGAAATGCCACAAAGAACTGAATCCC 1523
RESULT 34
PCT-US95-11878-1/c
; Sequence 1, Application PC/TUS9511878
; GENERAL INFORMATION:
; APPLICANT: Immunobiology Research, Institute Inc.
; TITLE OF INVENTION: Method for the Diagnosis of
; TITLE OF INVENTION: Depression Based on Monitoring Blood Levels of
; TITLE OF INVENTION: Arginine Vasopressin and/or Thymopoietin
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: 321 Norristown Road, Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version#1.25
; CURRENT APPLICATION DATA:
; FILING DATE: PCT/US95/11878
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,420
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: IRI46PCT
; TELEPHONE: (215) 540-9206
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2490 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 205..2286
PCT-US95-11878-1
Alignment Scores:
Pred. No.: 30.7 Length: 2490
Score: 71.50 Matches: 30
Percent Similarity: 43.24% Conservative: 18
Best Local Similarity: 27.03% Mismatches: 41
Query Match: 10.58% Indels: 22
DB: 5 Gaps: 4
US-10-048-196-2 (1-134) x PCT-US95-11878-1 (1-2490)
QY 25 GluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle-----40

Db 1838 GAAATAAACCTTAGGATTTTCATCATCTACTTTTGGAACTGATAATATCTGTGTGCTCC 1779
QY 41 -----LysHisAlaValLysGlnThrCys----- 48
Db 1778 CTGGAAGAACAATGGATATTTCATGTCAGTGAAGGAGCTGCTGTCAACTTGTCTGCCAT 1719
QY 49 GlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGlu 68
Db 1718 TCTTCTCAATTA---TTTAAATAGATTTCATGGAAGGAAAGATAGT 1671
QY 69 SerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAla 88
Db 1670 TCTGAGGATCTATTTTA---TCATGCTGTGATTGCTTAGCCACTTCTAGACCTTAAGGTA 1614
QY 89 ValSerLeuThrGluLeuThrThrAlaAla-IleAsnProAsnAlaArgThrGluValAla 108
Db 1613 GTGAGTGAATGAGACACAGCTTTTGGCAAAAGAGACATCAGTTCCGGATCCAGGTATG 1554
QY 108 aglnLysIleValArgHisSerLeuLysPro 118
Db 1553 TTCTGAAATGCCACAAAGAACTGAATCCC 1523
RESULT 35
US-09-010-809-14/c
; Sequence 14, Application US/09010809B
; Patent No. 6090601
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; TITLE OF INVENTION: Epithelone Polyketide Synthases and Encoding DNA
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: 30062-20020.00
; CURRENT APPLICATION NUMBER: US/09/010,809B
; CURRENT FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 14
; LENGTH: 794
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; FEATURE:
; OTHER INFORMATION: "n" at various positions throughout the sequence may
; OTHER INFORMATION: be A, T, C, G, other or unknown
US-09-010-809-14
Alignment Scores:
Pred. No.: 8.2 Length: 794
Score: 70.00 Matches: 28
Percent Similarity: 35.34% Conservative: 13
Best Local Similarity: 24.14% Mismatches: 43
Query Match: 10.36% Indels: 32
DB: 3 Gaps: 4
US-10-048-196-2 (1-134) x US-09-010-809-14 (1-794)
QY 17 LeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThr 36
Db 777 ATCGCTCCACACCCAGCGCTTCGCCAGTCGCCCTCGCCCTCAGCGCAGGACACC 718
QY 37 ThrSerLeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGln 56
Db 717 ACCAGC-----GAGTCCACCTTCGAGCGCGCTCCACGAGGCTCAGGAC-----673
QY 57 TyrTrpLys-----IleAlaAlaMetLysLeu 65
Db 672 ---TGGAGTCCACCCGGTACAGGTCTCGGCTCGGATGCTCGCTCCGCCAGCTGC 616
QY 66 SerSerGluSerLysAlaLys-----72
Db 615 TCGGGCGTCCCGCGCGAGATGACAGGACCGACCGCTCACCACCGGTCGGCGCGGCA 556
QY 73 -----IleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSer 90

DB: 3 4 Gaps: 4
US-10-048-196-2 (1-134) x US-08-776-265-2 (1-4565)
Qy 13 ThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSer 32
Db 1165 ACATTTTCGCAAAATTCGTTGTTGCTGCTCCCTCGAGCGTCAATTCGAGCACTTCT 1224
Qy 33 AlaAsnLeuThrThrSerLeuLeuLysHisAlaValLysGln----- 46
Db 1225 ACGAGCAAGTCCATCTTTTGTTCGCAATTCCTGTGAAACGATAGACCATTTAATTGGG 1284
Qy 47 ThrCysGlnThrGlnLeuThrGlyHisGlnThrProLysLeuAlaA----- 62
Db 1285 ACGTGTTCATTTTGTGACTTTTTTACTTAACATAAATCAGCTTCCACTAAGACGCC 1344
Qy 62 ----- 62
Db 1345 TCCAAAGTCGTTTGAACCTGCTGGCGTTTAAGTTCGCTATTAATGTAATGCACCAT 1404
Qy 63 -----MetLysLeuSerSerGluSerLysAlaLysLeuSerGluThrAlaCysGly 79
Db 1405 CCATTAATATGCTCTCTTCGTCGAGATCTCTATAAAATTTTACTTCTTTAATTATG 1464
Qy 80 CysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAla 99
Db 1465 TTAATAGGTGTGTTAGCCCGCTCCGCA---ACCATGACTGAATTTTAACCTCTCATCAT 1521
Qy 100 AsnProAsnAlaArgThrGluValAlaGlnLysLeuValArgHisSerLeuLysProCys 119
Db 1522 AAT----- 1524
Qy 120 MetLeuGluThrValAsnAlaPheLeuValProThrThrThr 133
Db 1525 -----ATTAGCACATAATAATACTAGTACTAGGACT 1554

RESULT 40
US-08-151-574-31/c
; Sequence 31, Application US/08151574
; Patent No. 5436156
; GENERAL INFORMATION:
; APPLICANT: Robert F.M. Van Gorcom
; APPLICANT: Willem Van Hartingsveldt
; APPLICANT: Petrus A. Van Paridon
; APPLICANT: Annemarie E. Veenstra
; APPLICANT: Rudolf G.M. Luttin
; TITLE OF INVENTION: Cloning and Expression of Microbial
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 545 Middlefield Road, Suite 200
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025-3471
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/151.574
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/688,578
; FILING DATE: 24-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 24615-20026.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-327-7250
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 6756 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Aspergillus ficuum (Aspergillus niger)
STRAIN: NRRL 3135
IMMEDIATE SOURCE:
LIBRARY: lambda AF
CLONE: pAF2-3, pAF2-6, pAF2-7
FEATURE:
NAME/KEY: exon
LOCATION: 210..253
FEATURE:
NAME/KEY: intron
LOCATION: 254..355
FEATURE:
NAME/KEY: exon
LOCATION: 356..1715
FEATURE:
NAME/KEY: CDS
LOCATION: join(210..253, 356..1715)
OTHER INFORMATION: /codon_start= 210
OTHER INFORMATION: /product= "phytase"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 210..380
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 381..1712
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "inositol phosphate
OTHER INFORMATION: /phosphatase"
OTHER INFORMATION: /product= "phytase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-151-574-31
Alignment Scores:
Pred. No.: 260 Length: 6756
Score: 69.50 Matches: 30
Percent Similarity: 42.24% Conservative: 19
Best Local Similarity: 25.86% Mismatches: 38
Query Match: 10.28% Indels: 29
DB: 1 Gaps: 6
US-10-048-196-2 (1-134) x US-08-151-574-31 (1-6756)
Qy 17 LeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThr 36
Db 6220 CTCGAAAGTCTGCAATCAACGCCGCAAGCCAGGCCG-----TGGATG 6179
Qy 37 ThrSerLeuLeuLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGln 56
Db 6178 ACGTTCTCTTGAACACCTTCTTTCCCGA-----AGACGCCCTTTGGCCCG----- 6131
Qy 57 TyrTrpLysLeuAlaAlaMetLysLeuSerSerGluSerLysAlaLysLysLeuSerGluThr 76
Db 6130 --TGGTTGCTGACAGCATTTGAAAGTACCCCTGCTCTTGACAGCAAGGCTGAATTCGACA 6074
Qy 77 AlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThr 96
Db 6073 CCATGT-----CTGGATACG 6059
Qy 97 AlaAlaLeuAsnPro-----AsnAlaArgThrGluValAlaGlnLysLeuValArgHis 114
Db 97 AlaAlaLeuAsnPro-----AsnAlaArgThrGluValAlaGlnLysLeuValArgHis 114

Db 6058 ACTTGGATACACCCGGAATCGAACGCAGTACTCCGTACGGCGAAAGTTTTTCCTTTTGAAT 5999
QY 115 SerLeuLysProCysMetLeuGluThrValAsnAlaPheIleValPro 130
Db 5998 ACCATGCTGATATGTATG---GAAACAAATTTGGTCTTGATGGTTCCA 5954

Search completed: May 5, 2003, 23:20:00
Job time : 66 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 5, 2003, 23:20:04 ; Search time 90 Seconds
(without alignments)
1759.303 Million cell updates/sec

Title: US-10-048-196-2

Perfect score: 676

Sequence: 1 MMKILYVATLMTAFTLASC.....SLKPCMLETNAFIVPTTR 134

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10048196/runat_28042003.151442.5681/app_query.fasta_1.327
-DB=Published_Applications_NA_QFMT-fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10048196.ecgn.1.1.84.@runat_28042003.151442.5681
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published_Applications_NA:**
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	85.5	12.6	3823	9	US-10-114-170-132
C 2	84	12.4	921	10	US-09-815-242-4588
C 3	84	12.4	957	10	US-09-815-242-8609
C 4	84	12.4	1799	7	US-08-781-986A-447

5	79.5	11.8	2456	10	US-09-822-849A-489	Sequence 489, App
6	78.5	11.6	600	9	US-09-738-626-2345	Sequence 2345, App
7	78	11.5	980	9	US-10-152-661-524	Sequence 524, App
8	78	11.5	980	9	US-09-866-050A-524	Sequence 524, App
9	77	11.4	2324	9	US-10-152-661-587	Sequence 587, App
10	77	11.4	2324	9	US-09-866-050A-587	Sequence 587, App
c 11	76.5	11.3	1161	10	US-09-864-761-19149	Sequence 19149, A
12	76.5	11.3	1818	10	US-09-801-368-427	Sequence 427, App
13	76.5	11.3	2179	9	US-09-854-133-152	Sequence 152, App
14	76.5	11.3	2179	10	US-09-738-973-152	Sequence 152, App
15	74.5	11.0	1193	10	US-09-819-505-3	Sequence 3, Appli
c 16	74.5	11.0	30365	10	US-09-825-414-1	Sequence 1, Appli
c 17	73.5	10.9	7193	9	US-10-071-338-1	Sequence 1, Appli
18	73	10.8	1143	10	US-09-972-186A-1	Sequence 1, Appli
c 19	73	10.8	1317	10	US-09-815-242-9978	Sequence 9378, Ap
c 20	73	10.8	10429	9	US-09-764-891-7375	Sequence 7375, Ap
21	73	10.8	14187	9	US-10-114-170-121	Sequence 121, App
22	72	10.7	1875	9	US-09-938-842A-1071	Sequence 1071, Ap
23	72	10.7	7104	10	US-09-815-242-4580	Sequence 4580, Ap
24	72	10.7	7107	10	US-09-815-242-8291	Sequence 8291, Ap
25	71.5	10.6	454	9	US-09-918-995-28770	Sequence 28770, A
26	71.5	10.6	4106	9	US-10-102-806-313	Sequence 313, App
c 27	71	10.5	768	9	US-09-938-842A-812	Sequence 812, App
c 28	71	10.5	2267	10	US-09-964-824A-248	Sequence 248, App
c 29	71	10.5	2267	10	US-09-880-107-2249	Sequence 2249, App
c 30	71	10.5	143306	10	US-09-729-920-3	Sequence 3, Appli
31	70.5	10.4	455	10	US-09-867-550-61	Sequence 61, Appl
c 32	70.5	10.4	980	10	US-09-864-761-9250	Sequence 9250, Ap
33	70.5	10.4	1169	10	US-09-799-777-112	Sequence 112, App
34	70.5	10.4	1775	9	US-10-098-841-26	Sequence 26, Appl
35	70.5	10.4	3089	9	US-09-992-598-46	Sequence 46, Appl
36	70.5	10.4	3089	9	US-09-989-293A-46	Sequence 46, Appl
37	70.5	10.4	3089	9	US-09-989-735-46	Sequence 46, Appl
38	70.5	10.4	3089	9	US-09-990-444-46	Sequence 46, Appl
39	70.5	10.4	3089	9	US-09-989-730-46	Sequence 46, Appl
40	70.5	10.4	3089	9	US-09-990-436-46	Sequence 46, Appl
41	70.5	10.4	3089	9	US-09-991-181-46	Sequence 46, Appl
42	70.5	10.4	3089	9	US-09-993-687-46	Sequence 46, Appl
43	70.5	10.4	3089	9	US-09-989-734-46	Sequence 46, Appl
44	70.5	10.4	3089	9	US-09-997-653-46	Sequence 46, Appl
45	70.5	10.4	3089	9	US-10-174-590-61	Sequence 61, Appl

ALIGNMENTS

RESULT 1

US-10-114-170-132/C

; Sequence 132, Application US/10114170

; Publication No. US20030023075A1

; GENERAL INFORMATION:

; APPLICANT: Blattner, Frederick R.

; Burland, Valerie

; Perna, Nicole T.

; Plunkett, Guy

; Welch, Rod

; TITLE OF INVENTION: NO. US20030023075A1el Sequences of E. coli O157

; NUMBER OF SEQUENCES: 265

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles & Brady

; STREET: 1 South Pinckney Street

; CITY: Madison

; STATE: WI

; COUNTRY: US

; ZIP: 53701-2113

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 8.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/114,170

; FILING DATE: 01-Apr-2002

; CLASSIFICATION: <Unknown>

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/453,702
: FILING DATE: 03-DEC-1999
: APPLICATION NUMBER: 60/110,955
: FILING DATE: 04-DEC-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Seay, Nicholas J.
: REGISTRATION NUMBER: 27386
: REFERENCE/DOCKET NUMBER: 960296.95017
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (608) 251-5000
: TELEFAX: (608) 251-9166
: INFORMATION FOR SEQ ID NO: 132:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3823
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 132:
US-10-114-170-132

```

Alignment Scores:		
Pred. No.:	0.122	Length:
Score:	85.50	Matches:
Percent Similarity:	66.91%	Conservative:
Best Local Similarity:	22.15%	Mismatches:
Query Match:	12.65%	Indels:
DB:	9	Gaps:

US-10-048-196-2 (1-134) x US-10-114-170-132 (1-3823)

Qy	6	TyrValThrAlaThrLeuMetThrAlaPhe--ThrLeuAlaSerCysAlaSerThrPro	24
		: : : : :	
Db	1462	TGGGTGACAAAGTAAGTGGGTGAGTTTTTGGTAACTCTCGGAGTGACATCAGAAACA	1403
		: : : : :	
Qy	25	GlusErAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaVal	44
		: : : : :	
Db	1402	GAAGGAACGTGGAAAGACATCGGAAGATCAACACACACGCGTATTTTCTTTTCTCTTGGTC	1343
		: : : : :	
Qy	45	LysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLys	64
		: : : : :	
Db	1342	ATCCAGCGCGGCGAACAACAGCAGCTCCGGCAAC-----	1310
		: : : : :	
Qy	65	LeuSerSerGluSerLysAlaLysIleSerGlnThrAlaCysGlyCysValAlaAspLys	84
		: : : : :	
Db	1309	ATATCCACAGTGGCATATTCACATCTTCACGACACTGCTCGGACGACATTCAGATAAG	1250
		: : : : :	
Qy	85	AlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArg	104
		: : : : :	
Db	1249	ATGCCTGCGCGCTTACGATTACCACGGGAATGACGTCGCATACCTACTCTGGCGTGGCG	1190
		: : : : :	
Qy	105	ThrGluValAlaGlnLysIleValArgHisSerLeuLysProCys-----	119
		: : : : :	
Db	1189	GTAATCAGTTTGACAAGCGCTTTAGCGCACTCTTATCTTCATGATTATGGCCACCAAGTA	1130
		: : : : :	
Qy	120	-----MetLeuGluThrValAsn	125
		: : : : :	
Db	1129	GCNTAGCCGCAGTATAGCCACCGGGGATGCTTTTGACTGGCGCTTGCTTAAGTATTCAGG	1070
		: : : : :	
Qy	126	AlaPheIleValProThrThrThrArg	134
		: : : : :	
Db	1069	TGCTTTTCTTCTTACCAATTACACCGAGG	1043
		: : : : :	

RESULT 2
US-09-815-242-4588/c
US-09-815-242-4588, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, kari L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Dantel

```

; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 4588
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4588

```

Alignment Scores:		
Pred. No.:	0.0216	Length:
Score:	84.00	Matches:
Percent Similarity:	45.30%	Conservative:
Best Local Similarity:	29.06%	Mismatches:
Query Match:	12.43%	Indels:
DB:	10	Gaps:
		921
		19
		44
		20
		6
US-10-048-196-2 (1-134)	x	US-09-815-242-4588 (1-921)

Qy	1	MetMetLysIleLeuYrValThrAlaThrLeuMetThrAla-----PheThr	16
Db	539	ATGCCACAGTACAGTACGTAACC--ACTTTTTTATCTGCAAAATAACGCTTTATTCTCT	483
Qy	17	LeuAlaSerCysAlaSerThr-----ProGluSer	26
Db	482	TTAATCCAGTCAGGCAAACTCTAAGACGCGTGATATTTGGACGAATTGCACCTCGAAA	423
Qy	27	AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGln	46
Db	422	TGACCTAAATCAAAATTCATTAATCATTAACGTGCATCAATAATGACTGTGTGTCATCATCTTCA	353
Qy	47	ThrCysGlnThrGlnLeuThrGlyHisGlnYrTrpLysIleAlaAlaMet-----Lys	64
Db	362	AGAGCTTTTCTAAATTCACAGGTGATAAATATTTGGCCAGTTGTGTCTTGGATCGACG	303
Qy	65	LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLys	84
Db	302	TCATTCTCTAAATCTAAGCAACAAATCTCTTTTCGAGGACGTACATGCATTTTCTTAAAA	243
Qy	85	Ala---ProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn	100
Db	242	GCATGTCTTCAGGTTCATCAAT-----TTAAACACCATATCTTTGAAAT	198

RESULT 3
US - 09-815-242-8609/c
: Sequence 8609, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Hasebebeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: zyskind, Judith W

APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8609
LENGTH: 957
TYPE: DNA
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(957)
US-09-815-242-8609

Alignment Scores:
Pred. No.: 0.0229 Length: 957
Score: 84.00 Matches: 34
Percent Similarity: 45.30% Conservative: 19
Best Local Similarity: 29.06% Mismatches: 44
Query Match: 12.43% Indels: 20
DB: 10 Gaps: 6

US-10-048-196-2 (1-134) x US-09-815-242-8609 (1-957)

Qy 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAla-----PheThr 16
Db 542 ATGCCACGAGTACAGTACGTAACC---ACTTTTATCTGCAAAATAACGCTTTATTCTCT 486
Qy 17 LeuAlaSerCysAlaSerThr-----ProGluSer 26
Db 485 TTAATCCAGTCAGGCAAAATCTCTAAAGCGGTGATATTGGACGAATTGCACCTCGGAAA 426
Qy 27 AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGln 46
Db 425 TGACCTAAATCAATCAATATCATCTAGTCATCAATAATGACTGTCTCATCTCTCA 366
Qy 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMet-----Lys 64
Db 365 AGAGCTTTTCTAAATTCATACAGGTGATAAATATTGGCCAGTTGTGTGCTTGGATCGAGC 306
Qy 65 LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLys 84
Db 305 TCTTCTTCTAAATCTAAAGCAACATCTCTTTTCGAGGACGTACATGTCATTTCTTAA 246
Qy 85 Ala---ProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
Db 245 GCATGCTCTTCAGCTTCATCAATT-----TTAAACACCATATCTTTGAAT 201

RESULT 4

US-08-781-986A-447/c

Sequence 447, Application US/08/781986A

Publication No. US20030054436A1

GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 447:
SEQUENCE CHARACTERISTICS:
LENGTH: 1799 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-447

Alignment Scores:
Pred. No.: 0.0617 Length: 1799
Score: 84.00 Matches: 34
Percent Similarity: 45.30% Conservative: 19
Best Local Similarity: 29.06% Mismatches: 44
Query Match: 12.43% Indels: 20
DB: 7 Gaps: 6

US-10-048-196-2 (1-134) x US-08-781-986A-447 (1-1799)

Qy 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAla-----PheThr 16
Db 663 ATGCCACGAGTACAGTACGTAACC---ACTTTTATCTGCAAAATAACGCTTTATTCTCT 607
Qy 17 LeuAlaSerCysAlaSerThr-----ProGluSer 26
Db 606 TTAATCCAGTCAGGCAAAATCTCTAAACGCTGATATTGGACGAATTGCACCTCGGAAA 547
Qy 27 AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGln 46
Db 546 TGACCTAAATCAATCAATATCATCTAGTCATCAATAATGACTGTGTCATCATCTTCA 487
Qy 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMet-----Lys 64
Db 486 AGAGCTTTTCTAAATTCATACAGGTGATAAATATTGGCCAGTTGTGTGCTTGGATCGAGC 427
Qy 65 LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLys 84
Db 426 TCATCTTCTAAATCTAAAGCAACGATTTCTTTTCGAGGACGTACATGTCATTTCTTAA 367
Qy 85 Ala---ProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
Db 366 GCATGCTCTTCAGCTTCATCAATT-----TTAAACACCATATCTTTGAAT 322

RESULT 5

US-09-822-849A-489
; Sequence 489, Application US/09822849A
; Patent No. US20020045170A1

GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.

; APPLICANT: Clark, Hilary

; APPLICANT: Fechtel, Kim

; APPLICANT: Agostino, Michael J.

; APPLICANT: Howes, Steven H.

; APPLICANT: Resnick, Richard J.

; APPLICANT: Gulukota, Kamalakara

; APPLICANT: Graham, James R.

; APPLICANT: Genetics Institute, Inc.

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS

; FILE REFERENCE: GIN 6403

; CURRENT APPLICATION NUMBER: US/09/822,849A

; PRIOR FILING DATE: 2001-09-04

; PRIOR APPLICATION NUMBER: 60/195,582

; PRIOR FILING DATE: 2000-04-06

; NUMBER OF SEQ ID NOS: 598

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 489

; LENGTH: 2456

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-822-849A-489

Alignment Scores:

Pred. No.: 0.455 Length: 2456

Score: 79.50 Matches: 39

Percent Similarity: 41.56% Conservative: 25

Best Local Similarity: 25.32% Mismatches: 48

Query Match: 11.76% Indels: 42

DB: 10 Gaps: 5

US-10-048-196-2 (1-134) x US-09-822-849A-489 (1-2456)

QY 1 MetMetLysLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20

Db 1756 TTGAGGAAGTCTCTTCTCTCAGAGCTAGCTTAAATGTATACAAAGCTGTCAAGCTGT 1815

QY 21 AlaSerThrProGluSer-----Ser-LeuIleLeu 26

Db 1816 GGTCATCGCCAGAGAGCGGTTTCACACAGCTGCAGATCTGATGTCGACCAACGAA 1875

QY 27 -----AsnProLysAsnSerSerAlaAsnLeuThr-----Ser-LeuIleLeu 41

Db 1876 TGAAGAAGTCCATGTGCGGTCAGTCTCTCTCACCAGAGGTTCTTCAATATCTTAT 1935

QY 41 sHisAlaValLys-----GlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrp 59

Db 1936 GCATAGCATCCAAAGTTAAAGGGTTGTGCACTAGCTCGAGAGAAATCAAGAATGGAA 1995

QY 59 sIle-AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCys 79

Db 1996 AATGCTGTGTAATGTCTGCTGACGCTACAGGAGAACTAGA----- 2036

QY 79 LysCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaI 99

Db 2037 -----ACATTAGAACTTTGGAAGAGGGCGGG 2064

QY 99 leAsnProAsnAlaArgThrGluValAlaGlnLysIleValArg-HisSerLeuLysPro 118

Db 2065 GACAATTGAATGATTTTGTTCACCTGCCAAGGTTGTTGCAGTCACTCATTTGAAAAAC 2124

QY 119 Cys---MetLeuCluThrValAsnAlaPheIleVal 129

Db 2125 ATTTCTCTGCTCCAGACAGAAAAAACTTTATAGTT 2160

RESULT 6

US-09-738-626-2345

; Sequence 2345, Application US/09738626

; Publication No. US20020197605A1

GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; PRIOR FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO 2345

; LENGTH: 600

; TYPE: DNA

; ORGANISM: Corynebacterium glutamicum

US-09-738-626-2345

Alignment Scores:

Pred. No.: 0.0695 Length: 600

Score: 78.50 Matches: 27

Percent Similarity: 47.83% Conservative: 28

Best Local Similarity: 23.48% Mismatches: 47

Query Match: 11.61% Indels: 13

DB: 9 Gaps: 4

US-10-048-196-2 (1-134) x US-09-738-626-2345 (1-600)

QY 1 MetMetLysLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20

Db 7 CTTTTCAGGCAACCCAGCTTACTTTTACTGTTGCAGCAGCATTTGGCGTGCAGCGTGC 66

QY 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrSerLeuIle 40

Db 67 TCCAGCAGTGTATGATTCCTCTTCGGAGTCAAGCACCTCTTCTTCCACCTCTTCGGCTGC 126

QY 41 LysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIle 60

Db 127 TCTGATGCTCCGACTCAGTACCCCACTGCTGAGGAAGTGAAT----- 168

QY 61 AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGlu-----ThrAla 77

Db 169 GCATTTTGGCTGTGCAACCCAGCTGAGCACCACCAATCGAGGAGAGAGTGAAGACTGTT 228

QY 78 CysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAla 97

Db 229 CAGGGT-----TCTGAGAACGCTCTCTGAGCTG-----TTTCAGACCACTGACTCAGGCA 276

QY 98 AlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleVal 112

Db 277 AAGGTGGAGTGGGGTGTCTGAGTTCAGGTTGTTGGTCTGTGCTT 321

RESULT 7

US-10-152-661-524

; Sequence 524, Application US/10152661

; Publication No. US20030022835A1

GENERAL INFORMATION:

; APPLICANT: Watson, James D.

; APPLICANT: Strachan, Lorna

; APPLICANT: Sleeman, Matthew

; APPLICANT: Onrust, Rene

; APPLICANT: Murison, James G.

APPLICANT: Kumble, Krishanand D.
TITLE OF INVENTION: Compositions Isolated From Skin Cells
FILE REFERENCE: 11000.1011c5
CURRENT APPLICATION NUMBER: US/10/152,661
CURRENT FILING DATE: 2002-05-20
PRIOR APPLICATION NUMBER: 09/866,050
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 60/221,232
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: 60/206,650
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 09/312,283
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: PCT/NZ99/00051
PRIOR FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 09/188,930
PRIOR FILING DATE: 1998-11-09
PRIOR APPLICATION NUMBER: 09/069,726
PRIOR FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 725
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 524
LENGTH: 980
TYPE: DNA
ORGANISM: Rat
US-10-152-661-524

Alignment Scores:
Pred. No.: 0.178 Length: 980
Score: 78.00 Matches: 33
Percent Similarity: 43.44% Conservative: 20
Best Local Similarity: 27.05% Mismatches: 53
Query Match: 11.54% Indels: 16
DB: 9 Gaps: 6

US-10-048-196-2 (1-134) x US-10-152-661-524 (1-980)

Qy 19 SerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSer 38
Db 44 TCCTGTGGCGCGCGCTGCAGCAATCCGAGCAAGACTTCT-----ACGACTTCA 94
Qy 39 LeuLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr--- 57
Db 95 AGCGGTGAATATCCGGGGCAAGCTGCTGCTGGAGAGTACCGCGCTCGTTTCCC 154
Qy 58 ---TrpLysIleAlaAlaMetLysLeuSerSerGluSerAlaLysIleSerGluThr 76
Db 155 TGGTGGTGAATATCCGGGGCAAGCTGCTGCTGGAGAGTACCGCGCTCGTTTCCC 154
Qy 77 AlaCysGlyCysValAlaAspLysAlaProGluAlaValSer---LeuThrGluLeuThr 95
Db 209 TGCAGCAGCTGCAGCGGACCTGGCCCTACCATTTTAACGTGCTTCCCTTCCA 268
Qy 96 ThrAlaAlaIleAsnProAsn-----AlaArgThrGluValAlaGln 109
Db 269 ACCAGTTTGGCCACACAGGAGACAGACAGCAACAGGAGATTGAGAACTTGGCCGCCCA 328
Qy 110 LysIleValArgHisSerLeuLysProCysMetLeuGluThrValAlaPheIleVal 129
Db 329 CCTACAGTG-----TCTCTTCCCATGTTTAGCAAGATCGCAGTCACTGCGACTGGTG 382
Qy 130 ProThr 131
Db 383 CCCACC 388

RESULT 8

US-09-866-050A-524
Sequence 524, Application US/09866050A
Publication No. US2003004071A1
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna

APPLICANT: Sleeman, Matthew
APPLICANT: Onrust, Rene
APPLICANT: Murison, James G.
APPLICANT: Kumble, Krishanand D.
TITLE OF INVENTION: Compositions Isolated From Skin Cells
FILE REFERENCE: 11000.1011c4U
CURRENT APPLICATION NUMBER: US/09/866,050A
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 725
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 524
LENGTH: 980
TYPE: DNA
ORGANISM: Rat
US-09-866-050A-524

Alignment Scores:
Pred. No.: 0.178 Length: 980
Score: 78.00 Matches: 33
Percent Similarity: 43.44% Conservative: 20
Best Local Similarity: 27.05% Mismatches: 53
Query Match: 11.54% Indels: 16
DB: 9 Gaps: 6

US-10-048-196-2 (1-134) x US-09-866-050A-524 (1-980)

Qy 19 SerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSer 38
Db 44 TCCTGTGGCGCGCGCTGCAGCAATCCGAGCAAGACTTCT-----ACGACTTCA 94
Qy 39 LeuLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr--- 57
Db 95 AGCGGTGAATATCCGGGGCAAGCTGCTGCTGGAGAGTACCGCGCTCGTTTCCC 154
Qy 58 ---TrpLysIleAlaAlaMetLysLeuSerSerGluSerAlaLysIleSerGluThr 76
Db 155 TGGTGGTGAATATCCGGGGCAAGCTGCTGCTGGAGAGTACCGAGCAAGACT-----ACGAGCCT 208
Qy 77 AlaCysGlyCysValAlaAspLysAlaProGluAlaValSer---LeuThrGluLeuThr 95
Db 209 TGCAGCAGCTGCAGCGGACCTGGCCCTACCATTTTAACGTGCTTCCCTTCCA 268
Qy 96 ThrAlaAlaIleAsnProAsn-----AlaArgThrGluValAlaGln 109
Db 269 ACCAGTTTGGCCACACAGGAGACAGACAGCAACAGGAGATTGAGAACTTGGCCGCCCA 328
Qy 110 LysIleValArgHisSerLeuLysProCysMetLeuGluThrValAlaPheIleVal 129
Db 329 CCTACAGTG-----TCTCTTCCCATGTTTAGCAAGATCGCAGTCACTGCGACTGGTG 382
Qy 130 ProThr 131
Db 383 CCCACC 388

RESULT 9

US-10-152-661-587
Sequence 587, Application US/10152661
Publication No. US20030022835A1
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Onrust, Rene
APPLICANT: Murison, James G.
APPLICANT: Kumble, Krishanand D.
TITLE OF INVENTION: Compositions Isolated From Skin Cells
FILE REFERENCE: 11000.1011c5
CURRENT APPLICATION NUMBER: US/10/152,661
CURRENT FILING DATE: 2002-05-20
PRIOR APPLICATION NUMBER: 09/866,050
PRIOR FILING DATE: 2001-05-24


```
; PRIOR APPLICATION NUMBER: 60/221,232
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/206,650
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 09/312,283
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/N299/00051
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 09/188,930
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: 09/069,726
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 587
; LENGTH: 2324
; TYPE: DNA
; ORGANISM: Mouse
US-10-152-661-587

Alignment Scores:
Pred. No.: 0.964 Length: 2324
Score: 77.00 Matches: 27
Percent Similarity: 37.98% Conservative: 17
Best Local Similarity: 23.28% Mismatches: 36
Query Match: 11.39% Indels: 36
DB: Gaps: 6

US-10-048-196-2 (1-134) x US-10-152-661-587 (1-2324)
QY 12 MetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSer 31
Db 1934 GTGACTGCGACCGCAATTTCTGTTTCTTCAAGTCTACCA-----CCAAAGGACCTTC 1984
QY 32 SerAlaAsnLeuThrThrSerLeu-----IleLysHisAla----- 43
Db 1985 TGTTCAGGAGTACACCAATTTCTGTTAACTTCCAGAGGTACACACCTGAAAAAT 2044
QY 44 -----ValLysGlnThrCysGlnThrGlnLeuThr-----GlyHisGln 56
Db 2045 ACTTAGGAGCGGAGTACATGCAATCTGTCGGTAACATGAGGAAGTCTCAACCTCAGAC 2104
QY 57 TyrTrpLysIleAlaAlaMetLysLeuSerGluSerLysAlaLysIleSerGluThr 76
Db 2105 TCCTGGAAGCCTGCACCT-----TCCACAAACATTAAATCCAAGAGGTG 2149
QY 77 AlaCysGlyCysValAlaAspLys-----IleLysHisAla----- 84
Db 2150 GGTGCGCACTGGTGGAGACAGATGCTCCCTCCCGTGGCCCATGGCTCTCTTGGTCT 2209
QY 85 ----AlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIle 99
Db 2210 TCATGCGCTGAGGGTGGGGCTAACTGGTGTAGTCTTGGCTGCTG 2257

RESULT 10
US-09-866-050A-587
; Sequence 587, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 587
```

```
; LENGTH: 2324
; TYPE: DNA
; ORGANISM: Mouse
US-09-866-050A-587

Alignment Scores:
Pred. No.: 0.964 Length: 2324
Score: 77.00 Matches: 27
Percent Similarity: 37.93% Conservative: 17
Best Local Similarity: 23.28% Mismatches: 36
Query Match: 11.39% Indels: 36
DB: Gaps: 6

US-10-048-196-2 (1-134) x US-09-866-050A-587 (1-2324)
QY 12 MetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSer 31
Db 1934 GTGACTGCGACCGCAATTTCTGTTTCTTCAAGTCTACCA-----CCAAAGGACCTTC 1984
QY 32 SerAlaAsnLeuThrThrSerLeu-----IleLysHisAla----- 43
Db 1985 TGTTCAGGAGTACACCAATTTCTGTTAACTTCCAGAGGTACACACCTGAAAAAT 2044
QY 44 -----ValLysGlnThrCysGlnThrGlnLeuThr-----GlyHisGln 56
Db 2045 ACTTAGGAGCGGAGTACATGCAATCTGTCGGTAACATGAGGAAGTCTCAACCTCAGAC 2104
QY 57 TyrTrpLysIleAlaAlaMetLysLeuSerGluSerLysAlaLysIleSerGluThr 76
Db 2105 TCCTGGAAGCCTGCACCT-----TCCACAAACATTAAATCCAAGAGGTG 2149
QY 77 AlaCysGlyCysValAlaAspLys-----IleLysHisAla----- 84
Db 2150 GGTGCGCACTGGTGGAGACAGATGCTCCCTCCCGTGGCCCATGGCTCTCTTGGTCT 2209
QY 85 ----AlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIle 99
Db 2210 TCATGCGCTGAGGGTGGGGCTAACTGGTGTAGTCTTGGCTGCTG 2257

RESULT 11
US-09-864-761-19149/c
; Sequence 19149, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecomics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
```

```

RESULT 12
US-09-801-368-427
; Sequence 427, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amnr
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improvin
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801.36
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 427
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-427

```

Alignment scores:		
Pred. No.:	0.775	1818
Score:	76.50	Matches: 30
Percent Similarity:	40.30%	Conservative: 24
Best Local Similarity:	22.39%	Mismatches: 71
Query Match:	11.32%	Indels: 9
DB:	10	Gaps: 3

US-10-048-196-2 (1-134) X US-09-801-368-421 (1-1818)

Qy	7	ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer	26
Db	460	CTGACGTCACAGAGTAGACGCCATTGCACACCGCGCTCAACCAACACCATCAACT	519
Qy	27	Asn-----ProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLys	41
Db	520	GATATACATCGGCACGTGCCACAGCATCAACGACAAAGTTGCAACATCCATTCGCCACA	579
Qy	42	HisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTrpTrpLysIleAla	61
Db	580	AGCAGCAGCATCGTCAACCTCTACCACCACCAAGTACCTCATCGTCG-----ACTTCTACG	633
Qy	62	AlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysVal	81
Db	634	ACCCTAAGTCATCATCATCAACATCTACCACCACCAAGTACTACGTCGTCGAACCCCAT	693
Qy	82	AlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThr-----AlaAlaIle	99
Db	694	TCCAGGAGCATCATCATCATCATCATCAACTCCCAACCAACATCGTCAGGCCCAT	753
Qy	100	AsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCys	119
Db	754	TCTACAGCAGCAGCGTCGTCACACTTCCACTTCCAACTCAACAGTACAACATCGCCAACTCTCTCT	813
Qy	120	MetLeuGluThrValAsnAlaPheIleValProThrThrThr	133
Db	814	TCAGCACCTCAACAGCTCGTCTAATATACAAACCAACCAACGAGCAGACA	855

```

; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliott, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 152
; LENGTH: 2179
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-738-973-152

```

Alignment Scores:	1.03	Length:	2179
Pred. No.:	76.50	Matches:	39
Score:	40.91%	Conservative:	24
Percent Similarity:	25.32%	Mismatches:	49
Best Local Similarity:	11.32%	Indels:	42
Query Match:	10	Gaps:	5
DB:	10		

US-10-048-196-2 (1-134) x US-09-738-973-152 (1-2179)

Qy	1	MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys	20
Db	1426	TTTGGAGAACTCTCTCTCTCAGAGCTACGTTAAATGATAACAAGCTGTCAAGTGT	1485
Qy	21	AlaSerThrProGluSer	26
Db	1486	GGTCTGCCAGAGACGGTTTCACGAAGTCGAGATCTGATTGATGCTGAGCAACGA	1545
Qy	27	-----AsnProLysAsnSerSerAlaAsnLeuThrThr-----Ser-LeuIleLeY	41
Db	1546	TGAAGAAGTCCATGTGGGTCAAGTTCTGTCTGCCTCACACGAGAGTCTTCAAACTAT	1605
Qy	41	sHisAlaValLys-----GlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrIlePly	59
Db	1606	GCATAGCATCCAAAGTTAAAGGGTTGTGCCACTAGCTCGAGAGGAATCAAGATGGAA	1665
Qy	59	sIle-AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysG	79
Db	1666	AATGTGTGTGAATTGGTCTGCAGTCTACAGGAGCAAGCTAGA-----	1706
Qy	79	LysValAlaIleAsnLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaI	99
Db	1707	-----ACATTAGAAGCTTTGGAGAGCGCGGG	1734
Qy	99	IleAsnProAsnAlaArgThrGluValAlaIleGlnLysIleValArg-HisSerLysPro	118
Db	1735	GAGAAATTGAATGATTTTGTTCCTCAACTGCCAAGGTGTGTTCAGTCACTCATTCGAAAC	1794
Qy	119	Cys---MetLeuGluThrValAsnAlaPheIleVal	129
Db	1795	ATTTTCTGCTCCAGACGAAAAAACTTTATAGTT	1830

```

RESULT 15
US-09-819-505-3
; Sequence 3, Application US/09819505
; Patent No. US20020142411A1
; GENERAL INFORMATION:
; APPLICANT: Paigen, Beverly
; Beier, David R.
; TITLE OF INVENTION: Antioxid
; of Use: T

```

BEIER, DAVID R.
TITLE OF INVENTION: Antioxidant Protein 2, Gene and Methods
of Use Therefor


```

; SEQ ID NO 7375
; LENGTH: 10429
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7375

Alignment Scores:
Pred. No.: 39
Score: 73.00
Percent Similarity: 39.84%
Best Local Similarity: 26.02%
Query Match: 10.80%
DB: 9

US-10-048-196-2 (1-134) x US-09-764-891-7375 (1-10429)

Qy 4 IleLeuTyValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThr 23
Db 8473 GTCATGTTGCAGCGGCAGCATCCACCATCTCCCTGACACTGGGTCGTGCTCAGCACT 8411
Qy 24 ProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
Db 8413 TCCACATCCCTCAGTCATCTCCAAACCATTCCTCTGGAGGTAGTATTATTGTC 8354
Qy 41 -----LysHisAlaVal-----LysGlnThrCysGlnThrGlnLeuThr 53
Db 8353 CCCACTAGACATGTCACGTGCAGAGAAATTAAAGAAACGTCTCTAACACCCACGCTATCA 8294
Qy 54 GlyHisGlnTyTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLysIle 73
Db 8293 GGCAGCAGAGTCAGAAAGAGAGCCGGC-----AGCTCTAACTCCAAAGCTGGCATC 8243
Qy 74 SerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGlu 93
Db 8242 CTTCCACCACCTCCATGTTGCACACACAGCAGGGACCCCGACAGTGAGCTCTGTCT 8183
Qy 94 LeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArg 113
Db 8182 CTAACACCTTATCTCACCTCTCCCAAC-----AGGCTCAAGAGGAAGTACTGTCTCA 8132
Qy 114 HisSerLeu 116
Db 8131 TTTTCCCTA 8123

RESULT 21
US-10-114-170-121
; Sequence 121, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: NO. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Finckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 121:
SEQUENCE CHARACTERISTICS:
LENGTH: 14187
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

Alignment Scores:
Pred. No.: 63.2 Length: 14187
Score: 73.00 Matches: 35
Percent Similarity: 42.02% Conservative: 15
Best Local Similarity: 29.41% Mismatches: 45
Query Match: 10.80% Indels: 24
DB: 9 Gaps: 7

US-10-048-196-2 (1-134) x US-10-114-170-121 (1-14187)

QY 8 ThrAlaThrLeuMetThr---AlaPheThrLeu---AlaSerCysAlaSerThrProGlu 25
Db 6591 ACAGCAACATGATCATACCTGGTTTCACACCTGGAGCATCTCTGCTTCAACC---GAA 6647
QY 26 SerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLeuHisAlaValLys 45
Db 6648 TCACAAATGCTTTTCTTCTGATTCCTTTACCAACAAAGCTCATC-----TTAATA 6698
QY 46 GluThrCysGlnThrGlnLeuThrGlyHis-----GlnTyrTrpLysIleAlaAla 62
Db 6699 CCAAGTGTGCTGCTGCCATTCGACAACTGACGACATGAGTATTTGGACTCTGTGTGA 6758
QY 63 MeLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAla 82
Db 6759 ATCAGAAATCTTCG-----GGTGTCTCTCTGTGT--- 6788
QY 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102
Db 6789 -----AACATCAACCATTTGCATTTCTAACATAATCTTTTGCATGTCCTCCAA 6833
QY 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeu 121
Db 6834 TCTCGTAACGAATCCATATTCCTTAATACAAACATGATCCAAAGCTTGTGCAATA 6890

RESULT 22
US-09-938-842A-1071
Sequence 1071, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1071
LENGTH: 1875
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1071
Alignment Scores:
Pred. No.: 3.68 Length: 1875
Score: 72.00 Matches: 24
Percent Similarity: 46.59% Conservative: 17
Best Local Similarity: 27.27% Mismatches: 27
Query Match: 10.65% Indels: 20
DB: 9 Gaps: 3
US-10-048-196-2 (1-134) x US-09-938-842A-1071 (1-1875)
QY 11 LeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsn 30
Db 251 CTAATCAGCGCTTACACGCATCATCTGATTCTCTCCG-----CCGCGAAAC 301
QY 31 SerSerAlaAsnLeuThrThrSerLeuLeuLeuHisAlaValLysGlnThrCysGlnThr 50
Db 302 TCCTCTGCTTCCCTGAGCTCGCGCTCATTTCC----- 334
QY 51 GluLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGlu---Ser 69
Db 335 -----TGGAAACACACGCGATGACGTCACCTTCGCGCTTACA 373
QY 70 LysAlaLysIleSerGluThrAlaCysGlyCysValAlaAlaAspLysAlaProGluAlaVal 89
Db 374 GCGCAAGAACTTCACAACTACGCGTCCGACCGTCTCTCCGAGCAGACTCATTTCAAAA 433
QY 90 SerLeuThrGluLeuThrThrAla 97
Db 434 ACTATTCCGCGGTGACACATCG 457

RESULT 23
US-09-815-242-4580
Sequence 4580, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0


```
; SEQ ID NO 4580
; LENGTH: 7104
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4580

Alignment Scores:
Pred. No.: 29.8      Length: 7104
Score: 72.00        Matches: 33
Percent Similarity: 39.64%      Conservative: 11
Best Local Similarity: 29.73%    Mismatches: 33
Query Match: 10.65%             Indels: 34
DB: 10                  Gaps: 5

US-10-048-196-2 (1-134) x US-09-815-242-4580 (1-7104)
Qy 9 AlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro 28
Db 6016 GCAACAGGTGAAACACACAGCACATCAGCAACAGTTTCAGCACTGATGCAATGAT 6075
Qy 29 Lys-----AsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAla 43
Db 6076 AAACCGCAAGCTAATAACAATTCCTAGTCAGATACATCAACAAC----- 6120
Qy 44 ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMet 63
Db 6121 ---AGTCGCAATGGATAATGATGTAAC----- 6147
Qy 64 LysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysVal----- 81
Db 6148 -----AGCAACAGAGTAGTAGTCACTAATAACGCCACCACTGATAAA 6192
Qy 82 -----AlaAspLysAlaProGluAlaValSerLeuThr-----GluLeuThr 95
Db 6193 CCTGCAACAGAGCAGATAATGCAACGCCAGCAGAAAGCGCAACAATAACAATAGTACA 6252
Qy 96 ThrAlaAlaIleAsnProAsnAlaArgThrGlu 106
Db 6253 ACTACAGCAACAACAGAGATGCAACACAGAA 6285

RESULT 24
US-09-815-242-8291
; Sequence 8291, Application US/09815242
; Patent No. US20020061369A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8291
; LENGTH: 7107
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7107)
US-09-815-242-8291

Alignment Scores:
Pred. No.: 29.8      Length: 7107
Score: 72.00        Matches: 33
Percent Similarity: 39.64%      Conservative: 11
Best Local Similarity: 29.73%    Mismatches: 33
Query Match: 10.65%             Indels: 34
DB: 10                  Gaps: 5

US-10-048-196-2 (1-134) x US-09-815-242-8291 (1-7107)
Qy 9 AlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro 28
Db 6016 GCAACAGGTGAAACACACAGCACATCAGCAACAGTTTCAGCACTGATGCAATGAT 6075
Qy 29 Lys-----AsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAla 43
Db 6076 AAACCGCAAGCTAATAACAATTCCTAGTCAGATACATCAACAAC----- 6120
Qy 44 ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMet 63
Db 6121 ---AGTCGCAATGGATAATGATGTAAC----- 6147
Qy 64 LysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysVal----- 81
Db 6148 -----AGCAACAGAGTAGTAGTCACTAATAACGCCACCACTGATAAA 6192
Qy 82 -----AlaAspLysAlaProGluAlaValSerLeuThr-----GluLeuThr 95
Db 6193 CCTGCAACAGAGCAGATAATGCAACGCCAGCAGAAAGCGCAACAATAACAATAGTACA 6252
Qy 96 ThrAlaAlaIleAsnProAsnAlaArgThrGlu 106
Db 6253 ACTACAGCAACAACAGAGATGCAACACAGAA 6285

RESULT 25
US-09-918-995-28770
; Sequence 28770, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 28770
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(454)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-28770

Alignment Scores:
Pred. No.: 0.468      Length: 454
Score: 71.50         Matches: 29
Percent Similarity: 40.00%      Conservative: 11
```

Best Local Similarity: 29.00% Mismatches: 43
Query Match: 10.58% Indels: 17
DB: 9 Gaps: 4

US-10-048-196-2 (1-134) x US-09-918-995-28770 (1-454)

QY 6 TyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGlu 25
DB 139 TGGGAGCGGCTCGCGCATCGAATGATATCTTCAGCGAGT-----TCCACCCACGC 192
QY 26 SerAsnPro-----LysAsnSerSerAlaAsn-----LeuThr 36
DB 193 TGGGACCCCAAGATACCTATACAGTCCCTGGAAGACTTCATCCCGAGAGCTGTTTGACA 252
QY 37 ThrSerLeuLeuLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGln 56
DB 253 CAGTCCCAAGTGTACATCATCACCACCAAGAGCTGCAGAACCAAGCTTATCATCTGTACAG 312
QY 57 TyrTrpLys-----lleAlaAlaMetLysLeuSerSerGluSerLys 70
DB 313 CTATGGAAGAAAGCTGATCGGCTGTCTGTGTCATCGAACACAAAGAGTACAGCGCA 372
QY 71 AlaLysIleSerGluThrAlaCysGlyCysValAlaAlaAspLysAlaProGluAlaValSer 90
DB 373 ATGCTCTCTCTTCAACCTGGGCTTGTGTGTATGCCAGCCCAAGACCTCGCCCTCG 432

RESULT 26

US-102-806-313

; Sequence 313, Application US/10102806

; Publication No. US20030054421A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA103PIC1

; CURRENT APPLICATION NUMBER: US/10/102,806

; PRIOR FILING DATE: 2002-03-22

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05881

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 846

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 313

; LENGTH: 4106

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (344)

; OTHER INFORMATION: n equals a,t,g, or c

US-102-806-313

Alignment Scores:

Pred. No.: 14.9 Length: 4106

Score: 71.50 Matches: 34

Percent Similarity: 38.73% Conservative: 21

Best Local Similarity: 23.94% Mismatches: 58

Query Match: 10.58% Indels: 29

DB: 9 Gaps: 5

US-10-048-196-2 (1-134) x US-10-102-806-313 (1-4106)

QY 9 AlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro 28

DB 738 GCTACTTCTTCAACAGCTGTATCT-----GCTTCAAGCCGACTGCCTCTCTCT 785

QY 29 LysAsnSerSerAlaAsn-----LysThrThrSerLeuIle 40

DB 786 TCAAGCATGTCAGCAACAATTGTACTGTGANTACGTATCATCAATTCCACAGCTTCAATG 845

QY 41 LysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTrpTrpLysIle 60

DB 846 AAG---GGTCTTACGACTACAGGAAACTCGTCTCTTAATAGCACATCTAACAATAAGTA 902
QY 61 AlaAlaMetLysLeuSerSerGluSerLysAla----- 71
DB 903 TCAGCAGTCCCTACAAATATATGGCTGCCAAGAAACATCTACCCCCAAAATAAATTTTGT 962
QY 72 -----LysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAla 88
DB 963 GGTGGTAATAAGCTGCAGTCAACAGGAATAAAGCAGAGACACAAAGGACCCGANTGT 1022
QY 89 ValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAla 108
DB 1023 GTTAAAGTACTCTGTCTACTTCTGTGTCAGATTCTCT-----GAAGTAAAG 1070
QY 109 GlnLysIleValArgHisSerLeuLysProCysMetLeuGluThrValAsnAlaPheIle 128
DB 1071 CAAGACACAGTGTGAGAACCCAGTCACTCTTGTGCTTTACAGAGTGATGTG 1130
QY 129 ValPro 130
DB 1131 CAGCCA 1136

RESULT 27

US-09-938-842A-812/c

; Sequence 812, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE REFERENCE: SCRIPI300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A

; PRIOR FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 812

; LENGTH: 768

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-812

Alignment Scores:

Pred. No.: 1.27 Length: 768

Score: 71.00 Matches: 24

Percent Similarity: 39.37% Conservative: 26

Best Local Similarity: 18.90% Mismatches: 53

Query Match: 10.50% Indels: 24

DB: 9 Gaps: 2

US-10-048-196-2 (1-134) x US-09-938-842A-812 (1-768)

QY 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer 26

DB 720 GTGACCGTATCCAGAGCAGCAGCATACCTCCACCTCTCTCTCCACCTCCGACACC 661

QY 27 AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGln 46

DB 660 ACCACCGCTCTCTCTCTCCACCATGTCACCATGTC----- 622

QY 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTrpLysIleAlaAlaMetLysLeuSer 66

DB 621 -----AGCACCCACCCGCTATCTCTCTCC 598

QY 67 SerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaPro 86

```
Db 597 AGCTCCACCTCCTCCTCGTAACACACCATGAGCTCCTCCTCGCGAGAACCTCCTCC 538
Qy 87 GluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsnAlaAArgThrGlu 106
Db 537 TCCACCCCAACATTCACCGCCTCCACCACTCCATGTCCTCCCGGCGACCTCCACCTCC 478
Qy 107 ValAlaGlnLysIleValAArgHisSerLeuLysProCysMetLeuGluThrValAsnAla 126
Db 477 ATATCTCTCC-----TCCAGCACCAGCACCTTGTCCACCAACCGTATCTCCTCC 433
Qy 127 PheIleValProThrThrThr 133
Db 432 ACCATGTGCACCAACCAAC 412
RESULT 28
US-09-964-824A-248/c
; Sequence 248, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 248
; LENGTH: 2267
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-248
Alignment Scores:
Pred. No.: 6.93 Length: 2267
Score: 71.00 Matches: 28
Percent Similarity: 38.85% Conservative: 26
Best Local Similarity: 20.14% Mismatches: 50
Query Match: 10.50% Indels: 35
DB: 10 Gaps: 4
US-10-048-196-2 (1-134) x US-09-964-824A-248 (1-2267)
Qy 20 CysAlaSerThrPro-GluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLe 39
Db 1432 TGCGCCAAACATCTCTGCAGGGGAAGCCCATCAAAGCAGTGGCGCAGTGAGGCCACT 1373
Qy 39 uIleLysHis-----AlaValLysGlnThrCysGlnThrGlnLeuThrG1 54
Db 1372 GGTGTCTATAGTAGTCCACAGGTGCCTCTGAGCTGTACCTTCACCTGCAGGC 1313
Qy 54 yHisGlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSe 74
Db 1312 TCACCTGCCAGGGCCACTCTCCCAAGAGAGTTGTTCTCCCAACAATGCG----- 1261
Qy 74 rGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLe 94
Db 1260 ----TGTCCTTGTGTTTGTGTCGACAGAGAGTTGTCCTCCCAACAATCTCAAAG 1205
Qy 94 uThrThrAlaAla-----IleAsnProAsnAlaAArgThr----- 105
Db 1204 AGTAACCAAGAGCTCCCTGTGTCCTCCATACGCAATCTAGTTGGAGAACCATCCATAGATA 1145
Qy 106 -----GluValAlaG1 109
Db 1144 ATCTTAAGAAACACTTACACTTCTCTCTTACAGTCTTCTCTGGAGTAAGAATAAGTGA 1085
Qy 109 nLysIleValAArgHisSerLeuLysProCysMetLeuGluThrValAsnAlaPhe 127
Db 1084 AAAAATGACAGCAATCATCTTTGTGCAAGTCTCTTGGCAAAACATTCACCTCTTT 1030
RESULT 30
US-09-929-920-3/c
; Sequence 3, Application US/09729920
```

```
Qy 109 nLysIleValAArgHisSerLeuLysProCysMetLeuGluThrValAsnAlaPhe 127
Db 1084 AAAAATGACAGCAATCATCTTTGTGCAAGTCTCTTGGCAAAACATTCACCTCTTT 1030
RESULT 29
US-09-880-107-2249/c
; Sequence 2249, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2249
; LENGTH: 2267
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M13143
US-09-880-107-2249
Alignment Scores:
Pred. No.: 6.93 Length: 2267
Score: 71.00 Matches: 28
Percent Similarity: 38.85% Conservative: 26
Best Local Similarity: 20.14% Mismatches: 50
Query Match: 10.50% Indels: 35
DB: 10 Gaps: 4
US-10-048-196-2 (1-134) x US-09-880-107-2249 (1-2267)
Qy 20 CysAlaSerThrPro-GluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLe 39
Db 1432 TGCGCCAAACATCTCTGCAGGGGAAGCCCATCAAAGCAGTGGCGCAGTGAGGCCACT 1373
Qy 39 uIleLysHis-----AlaValLysGlnThrCysGlnThrGlnLeuThrG1 54
Db 1372 GGTGTCTATAGTAGTCCACAGGTGCCTCTGAGCTGTACCTTCACCTGCAGGC 1313
Qy 54 yHisGlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSe 74
Db 1312 TCACCTGCCAGGGCCACTCTCCCAAGAGAGTTGTTCTCCCAACAATGCG----- 1261
Qy 74 rGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLe 94
Db 1260 ----TGTCCTTGTGTTTGTGTCGACAGAGAGTTGTCCTCCCAACAATCTCAAAG 1205
Qy 94 uThrThrAlaAla-----IleAsnProAsnAlaAArgThr----- 105
Db 1204 AGTAACCAAGAGCTCCCTGTGTCCTCCATACGCAATCTAGTTGGAGAACCATCCATAGATA 1145
Qy 106 -----GluValAlaG1 109
Db 1144 ATCTTAAGAAACACTTACACTTCTCTCTTACAGTCTTCTCTGGAGTAAGAATAAGTGA 1085
Qy 109 nLysIleValAArgHisSerLeuLysProCysMetLeuGluThrValAsnAlaPhe 127
Db 1084 AAAAATGACAGCAATCATCTTTGTGCAAGTCTCTTGGCAAAACATTCACCTCTTT 1030
RESULT 30
US-09-929-920-3/c
; Sequence 3, Application US/09729920
```

```
; Patent No. US20020103115A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS.
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000858
; CURRENT APPLICATION NUMBER: US/09/729,920
; CURRENT FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 143306
; TYPE: DNA
; ORGANISM: Human
US-09-729-920-3

Alignment Scores:
Pred. No.: 4,68e+03 Length: 143306
Score: 71.00 Matches: 21
Percent Similarity: 43.48% Conservative: 19
Best Local Similarity: 22.83% Mismatches: 28
Query Match: 10.50% Indels: 24
DB: 10 Gaps: 3

US-10-048-196-2 (1-134) x US-09-729-920-3 (1-143306)
Qy 45 LysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyTrpLysIleAla----- 61
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 57563 NAAATCTCTGCGGTGATCTTGTGCGCACAGACATGCTGGGAGCAGCCCGGGAG 57504
Qy 62 -----AlaMetLysLeuSerSerGluSerLysLysLysIle 73
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 57503 CAACGTACCCGAGTCAGACCCAGCTGTGGAGATCAGGCGCAGCAGCCCTGATC 57444
Qy 74 SerGluThr-----AlaCysGlyCys----- 80
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 57443 ACCAACACCCCTCCAGTTAGAGGCTGTAGCGCATGCTCATGCCCAACAGCCCCAC 57384
Qy 81 ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 57383 TTCACAGACAAGAACTGAGACCCAGAGAGTTCAATCGCTAAGTCAGGTTAAGCGC 57324
Qy 101 ProAsnAlaArgThrGluValAlaGlnLysIleVal 112
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 57323 CCAAAAGCGAGAGCAGATTAAGGGAATCATC 57288

RESULT 31
US-09-867-550-61
; Sequence 61, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; PRIOR FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
```

```
; OTHER INFORMATION: wherein any n is one of a or t or g or c
US-09-867-550-61

Alignment Scores:
Pred. No.: 0.657 Length: 455
Score: 70.50 Matches: 29
Percent Similarity: 41.82% Conservative: 17
Best Local Similarity: 26.36% Mismatches: 38
Query Match: 10.43% Indels: 26
DB: 10 Gaps: 3

US-10-048-196-2 (1-134) x US-09-867-550-61 (1-455)
Qy 9 AlaThrLeuMetThrAlaPheThrLeuAlaSerCys----- 20
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 9 GCCTCGCTCATGACAGCGCGGATGCTTCGGGATTCGTCGATAGTCGGTCAAGCCGTCG 68
Qy 21 -----AlaSerThrProGluSerAsnProLysAsnSerSer 32
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 69 CCGGTGAGCGCGAGAGAGGGCGGCATCTCGCCGCTCTTCGAGGTGCGCCCTCGTCG 128
Qy 33 AlaAsnLeuThrThr-----SerLeuIleLysHisAlaValLysGlnThrCysGlnThr 50
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 129 GCGCGATCTCGACTTGTGCGCCCTCGTGAAGAGGGCGCAAGGTCAATCTGTGACGG 188
Qy 51 GlnLeuThrGlyHisGlnTyTrpLysIleAlaAlaMetLysLeuSerSerGluSerLys 70
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 189 CTACCTTCGACACCAACATCATCAG-----CAACGTAGTCAA 227
Qy 71 AlaLysIleSerGluThrAlaCysGlyCysVal-AlaAspLysAlaProGluAlaValSe 90
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 228 GGTGCGACCCCTCAGAGTCATTTGTTGCCCTGTTGGCGAGAGCAAGAGAGAGAGTCTGT 287
Qy 90 rLeuThrGluLeuThrThrAlaAlaIle 99
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
Db 288 CATCTCTCGCTGACGCGTGGCGCAATC 315

RESULT 32
US-09-864-761-9250/c
; Sequence 9250, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
```


	PRIOR APPLICATION NUMBER:	60/089553
	PRIOR FILING DATE:	1998-06-17
	PRIOR APPLICATION NUMBER:	60/089801
	PRIOR FILING DATE:	1998-06-18
	PRIOR APPLICATION NUMBER:	60/089907
	PRIOR FILING DATE:	1998-06-18
	PRIOR APPLICATION NUMBER:	60/089908
	PRIOR FILING DATE:	1998-06-18
	PRIOR APPLICATION NUMBER:	60/089947
	PRIOR FILING DATE:	1998-06-19
	PRIOR APPLICATION NUMBER:	60/089948
	PRIOR FILING DATE:	1998-06-19
	PRIOR APPLICATION NUMBER:	60/089952
	PRIOR FILING DATE:	1998-06-19
	PRIOR APPLICATION NUMBER:	60/090246
	PRIOR FILING DATE:	1998-06-22
	PRIOR APPLICATION NUMBER:	60/090252
	PRIOR FILING DATE:	1998-06-22
	PRIOR APPLICATION NUMBER:	60/090254
	PRIOR FILING DATE:	1998-06-22
	PRIOR APPLICATION NUMBER:	60/090349
	PRIOR FILING DATE:	1998-06-23
	PRIOR APPLICATION NUMBER:	60/090355
	PRIOR FILING DATE:	1998-06-23
	PRIOR APPLICATION NUMBER:	60/090429
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090431
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090435
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090444
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090445
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090472
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090535
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090540
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090542
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090557
	PRIOR FILING DATE:	1998-06-24
	PRIOR APPLICATION NUMBER:	60/090676
	PRIOR FILING DATE:	1998-06-25
	PRIOR APPLICATION NUMBER:	60/090678
	PRIOR FILING DATE:	1998-06-25
	PRIOR APPLICATION NUMBER:	60/090690
	PRIOR FILING DATE:	1998-06-25
	PRIOR APPLICATION NUMBER:	60/090694
	PRIOR FILING DATE:	1998-06-25
	PRIOR APPLICATION NUMBER:	60/090695
	PRIOR FILING DATE:	1998-06-25
	PRIOR APPLICATION NUMBER:	60/090696
	PRIOR FILING DATE:	1998-06-25
	PRIOR APPLICATION NUMBER:	60/090862
	PRIOR FILING DATE:	1998-06-26
	PRIOR APPLICATION NUMBER:	60/090863
	PRIOR FILING DATE:	1998-06-26
	PRIOR APPLICATION NUMBER:	60/091360
	PRIOR FILING DATE:	1998-07-01
	PRIOR APPLICATION NUMBER:	60/091478
	PRIOR FILING DATE:	1998-07-02
	PRIOR APPLICATION NUMBER:	60/091544
	PRIOR FILING DATE:	1998-07-01
	PRIOR APPLICATION NUMBER:	60/091519
	PRIOR FILING DATE:	1998-07-02
	PRIOR APPLICATION NUMBER:	60/091626
	PRIOR FILING DATE:	1998-07-02
	PRIOR APPLICATION NUMBER:	60/091633
	PRIOR FILING DATE:	1998-07-02
	PRIOR APPLICATION NUMBER:	60/091978


```

; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Alignment Scores:
Pred. No.: 13 3 Length: 3089
Score: 70.50 Matches: 36
Percent Similarity: 42.22% Conservative: 21
Best Local Similarity: 26.67% Mismatches: 56
Query Match: 10.43% Indels: 22
DB: 9 Gaps: 5

US-10-048-196-2 (1-134) x US-09-992-598-46 (1-3089)
Qy 10 ThrLeuMetThrAlaPheThr-----LeuAlaSerCysAlaSerThr 23
Db 726 ACCTTGATGGTGGTCTTGCATACGACGCGGCGGCTGCTGCTCCCTGGCTGCATCT 785
Qy 24 ProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAla 43
Db 786 -----ACTTCTTCACCCCAATGTCACAGCTGGGAATACAGG--- 824
Qy 44 ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnIleTyrTrpLysIleAlaAlaMet 63
Db 825 ---AGACGTACGTCGCACATATTGTCATACCAAGAGAGGCTGTCATCTGTGCGCGCAGC 881
Qy 64 LysLeuSerSerGluSer---LysAlaLysIle-----SerGluThr 76
Db 882 CCACGGAGCGAGCATCATGACAGCCAGAGTGTGATCTGGACCTGGCCCTACGGGGACC 941
Qy 77 AlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThr 96
Db 942 GCGTCTGGTGGCTCTTCAAGCGCCAGCGGAGAGCCCATCTACAGCAACGACTTCG 1001
Qy 97 AlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIle-ValArgHisSerLe 116
Db 1002 ACACCTACATCACCCTTCAGCGGCCACCTCATCAAGCGCGAGGAGCACTGAGGGCTCTGG 1061
Qy 116 uLysProCysMetLeuGluThrValAsnAlaPheIleValPro 130
Db 1062 GCCACCTCCCGCTGGAGAGCTCAGGTGCTGGTCCCTGCC 1104

RESULT 36
US-09-989-293A-46
; Sequence 46, Application US/09989293A
; Patent No. US20020177164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin

```


APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tomas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P27301C61
CURRENT APPLICATION NUMBER: US/09/989,735
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435

```

; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Alignment Scores:
Pred. No.:      13.3      Length:      3089
Score:           70.50    Matches:      36
Percent Similarity: 42.2%  Conservative: 21
Best Local Similarity: 26.6% Mismatches:    56
Query Match:      10.4%   Indels:       22
Db:                9      Gaps:         5

US-10-048-196-2 (1-134) x US-09-989-735-46 (1-3089)

Qy 10 ThrLeuMetThrAlaPheThr-----LeuAlaSerCysAlaSerThr 23
Db 726 ACCTGTGATGGGTGATGACCGGACGCGGACGTTGCTGCTCCCTGCGTGCATCT 785
Qy 24 ProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAla 43
Db 786 -----ACTTCTCAGCCTCAATGTGCACAGCTGGATTACAGG--- 824
Qy 44 ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMet 63
Db 825 ---AGACGTACGTGCACATATGATGACACGAAAGAGGCTGTCTCATCTGTACGCGCAGC 881
Qy 64 LysLeuSerSerGluSer---LysAlaLysIle-----SerGluThr 76
```

```

Db 882 CCAGCGAGCGCAGCATCATGCGACGACGAGGTGTGATGCTGGACCTAGCGGAGCC 941
Qy 77 AlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThr 96
Db 942 GCGTCTGGTGGGCTCTTCAAGCGCGGAGGAGAACGCCATCTACAGCAACGACTTCG 1001
Qy 97 AlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIle-ValArgHisSerLe 116
Db 1002 ACACCTATACATCACTTCAGCGGCCACCTCATCAAGGCGGAGGACGACTGAGGCGCTCTGG 1061
Qy 116 uLysProCysMetLeuGluThrValAsnAlaPheIleValPro 130
Db 1062 GCCACCTCTCCGCTGGAGAGCTCAGGTGCTGCTCCGTCCTCC 1104

RESULT 38
US-09-990-444-46
; Sequence 46, Application US/09990444
; Publication No. US20020193300A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC19
; CURRENT APPLICATION NUMBER: US/09/990,444
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
```

, PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907

Alignment Scores:

4

;
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26

;
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09
;
Alignment Scores:
Pred. No.: 13.3 Length: 3089
Score: 70.50 Matches: 36
Percent Similarity: 42.22% Conservative: 21
Best Local Similarity: 26.67% Mismatches: 56
Query Match: 10.43% Indels: 22
DB: 9 Gaps: 5
US-10-048-196-2 (1-134) x US-09-989-730-46 (1-3089)
Qy 10 ThrLeuMetThrAlaPheThr-----LeuAlaSerCysAlaSerThr 23
Db 726 ACCTGATGGTCTTTGACATGGCGACGGCCAGTTTGTCTCCCTCGGTGGCATCT 785
Qy 24 ProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLysHisAla 43
Db 786 -----ACTTCTTCAGCCTCAATGTGCACAGCTGGAATTACAAGG--- 824
Qy 44 ValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMet 63
Db 825 ---AGAGGTACGTGCACATTATGCATTAACCGAAGAGGCTGTCTACCTCGCGCAGC 881
Qy 64 LysLeuSerSerGluSer---LysAlaLysIle-----SerGluThr 76
Db 882 CCAGCGAGCGCAGCATCATGCAGCGCAGAGTGTGTATGTGTGGACCTGGCCTACGGGACC 941
Qy 77 AlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThr 96
Db 942 GCCTCTGGGTGGCGCTCTTCAAGCGCCAGCGCGAGAACGCCATCTACAGCAACGACTTCG 1001
Qy 97 AlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIle-ValArgHisSerLe 116
Db 1002 ACACCTACATCACCTTCAGGGGCCACCTCATCAAGGCCGAGGAGGAGGAGGAGGAGGAG 1061
Qy 116 uLysProCysMetLeuGluThrValAsnAlaPheIleValPro 130
Db 1062 GCCACCCCTCCCGCTGGAGAGCTCAGTGTGTGTCTCCCGTCCCC 1104
RESULT 40
US-09-990-436-46
; Sequence 46, Application US/09990436
; Publication NO. US20020198148A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C14
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US/09/990,436
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 5, 2003, 22:18:19 ; Search time 1060 Seconds
(without alignments)
2047.354 Million cell updates/sec

Title: US-10-048-196-2

Perfect score: 676

Sequence: 1 MKKILYVATLMTAFTLASC.....SLKPCMLETNVAIVPTTTR 134

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US10048196/runat_28042003_151440_5598/app_query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10048196_ECGN_1_1456_@runat_28042003_151440_5598 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c	1	98.5	14.6	476	17	AQ886059 HS_5530_B
	2	95	14.1	813	14	BQ839107 WHE3589_D
	3	94.5	14.0	709	14	BQ804850 WHE3559_G
	4	94.5	14.0	769	14	BQ804254 WHE3552_C
	5	94.5	14.0	784	14	BQ807264 WHE3588_G
	6	94	13.9	1080	14	BM802615 AGENCOURT
	7	93	13.8	679	17	AG158466 Pan trogl
	8	92.5	13.7	427	10	AW504179 UI-HF-BN0
	9	92	13.6	679	17	AZ358694 1M0101B17
	10	91	13.5	811	9	AJ452083 AJ452083
	11	90	13.3	630	14	BQ606191 BRY_2034
	12	90	13.3	787	14	BQ807035 WHE3586_A
	13	90	13.3	812	14	BQ804628 WHE3556_H
	14	89.5	13.2	610	13	BJ233642 BJ233642
	15	89.5	13.2	646	12	BF429171 WHE1709_C
	16	89.5	13.2	671	14	BQ246193 TAE15015A
	17	89.5	13.2	676	14	BQ245146 TAE15030B
	18	89.5	13.2	754	14	BQ246348 TAE15012H
	19	89	13.2	677	14	BQ245845 TAE15019D
	20	89	13.2	694	14	BQ246558 TAE15008E
	21	89	13.2	731	14	BQ246454 TAE15010E
	22	88.5	13.1	792	14	BQ839009 WHE3550_C
	23	88	13.0	607	12	BG262346 WHE0874_G
	24	88	13.0	708	14	BQ251408 TAE25029B
	25	88	13.0	714	13	BM634463 170006875
	26	87.5	12.9	808	14	BQ838664 WHE3591_C
	27	87.5	12.9	849	13	BJ237941 BJ237941
	28	87	12.9	536	10	BE423599 WHE0072_H
	29	87	12.9	631	14	BQ246932 TAE15003D
	30	87	12.9	631	14	BQ244975 TAE15032G
	31	87	12.9	641	10	AW448815 BRY_1500
	32	87	12.9	641	13	BI968881 GMB30006B
	33	87	12.9	641	14	BQ605913 BRY_1500
	34	87	12.9	643	14	BQ251877 TAE25023A
	35	87	12.9	647	14	BQ246079 TAE15016E
	36	87	12.9	652	14	BQ244638 TAE15039G
	37	87	12.9	652	14	BQ249968 TAE25011E
	38	87	12.9	657	14	BQ249079 TAE25048C
	39	87	12.9	658	14	BQ244737 TAE15038C
	40	87	12.9	674	14	BQ250428 TAE25005B
	41	87	12.9	675	10	AW310069 SF30e05.x
	42	87	12.9	695	14	BQ24732 TAE15038C
	43	87	12.9	705	14	BQ249449 TAE25043F
	44	87	12.9	707	14	BQ251394 TAE25029C
	45	87	12.9	720	14	BQ246266 TAE15014B

ALIGNMENTS

RESULT 1
AQ886059/c
LOCUS AQ886059 476 bp DNA linear GSS 09-NOV-1999
DEFINITION HS_5530_B2_C07_SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-9298 Col-14 Row-F, DNA sequence.
ACCESSION AQ886059
VERSION AQ886059.1 GI:6317526
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 476)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and

Db	140	GGGGAATTCCTTTTACACTAGTACCCACCAAGAGGGAAATGTTCAATTTTACACCTT	81
Qy	125	Asn 125 	
Db	80	AAT 78	
RESULT 2			
BQ839107			
LOCUS			
DEFINITION			
ACCESSION	BQ839107	813 bp mRNA	linear EST 08-AUG-2002
VERSION	BQ839107.1	WHE3589_D08_H15ZS Wheat developing grains	cdna library
KEYWORDS	EST.	WHE3589_D08_H15, mRNA	sequence.
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

FEATURES

source

Location/Qualifiers

1. 813

/organism="Triticum aestivum"

/cultivar="Butte 86"

/db_xref="taxon:4565"

/clone="WHE3589_D08_H15"

/clone_lib="Wheat developing grains cDNA library"

/tissue_type="whole grains"

/dev_stage="3-44 days post anthesis seed"

/lab_host="E. coli SOLR"

/note="Vector: Lambda ZAP II, excised phagemid: Site1:
ECORI; Plants were grown under six following different
environmental regimes in greenhouse, Environment 1)
240C/170C day/night, well-watered, with post-anthesis
fertilizer, Environment 2) 240C/170C day/night,
well-watered, without post-anthesis fertilizer,
Environment 3) 370C/170C day/night, well-watered, with
post-anthesis fertilizer, Environment 4) 370C/170C
day/night, well-watered, without post-anthesis fertilizer,
Environment 5) 370C/170C day/night plus drought, with
post-anthesis fertilizer, Environment 6) 370C/170C
day/night plus drought, without post-anthesis fertilizer,
developing wheat grains from the following were excised
and frozen in liquid nitrogen, Environment 1 at 3, 5, 7,
8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment
2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44
DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28,
32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24,
28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16,
20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16,
20, 24, 28, 30 DPA and total RNA was prepared by S.
Altendach and K. Cronin at USDA-ARS, Albany, CA. A cDNA
library was made using poly (A) RNA, and the cDNA clones
were in vivo excised to give pluescript SK(-) phagemids
in the T3 Close lab (Chin, Close, Fenton) at the
University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD

```

BASE COUNT      276 a 258 c 126 g 153 t
ORIGIN

Alignment Scores:
Pred. No.:      0 699      Length:      813
Score:          95.00      Matches:      36
Percent Similarity: 41.73%      Conservative: 22
Best Local Similarity: 25.90%      Mismatches: 47
Query Match:    14.05%      Indels:      34
DB:             14      Gaps:      6

US-10-048-196-2 (1-134) x B0839107 (1-813)

Qy 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer 26
Db 336 GTATCCGCAACCAACAAGCAATTCGCAGCAACAGCAACAACAACAACAACA 395
Qy 27 AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeLysHisAlaValLysGln 46
Db 396 AATCCTTCAACAATAATTCGCAACAACAACAACAACAACAACAACAACA 437
Qy 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTrpLysIleAlaAlaMet-----Lys 64
Db 438 GGATGCTGCTGTGCACAAACCAACATAGACATGCACAGCTCACAAGTATCTCAACAAG 497
Qy 65 LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLys 84
Db 498 TTACCATATTTGCAACAATAATTTGTCAGCAACT-----GTGCAGAC--- 542
Qy 85 AlaProGluAlaValSerLeu-----Lys 64
Db 543 ---CCCCAGCTGTGCACAGTGCACAGTCCCAAGTCATCCACAATGCTGCTTCATGCTATTATTCTGCA 599
Qy 92 -----ThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluVal 107
Db 600 TCATCAATAACAACAAAACAACAACAACAACCGTCGAGCGAGGCTCTCTACCAGCACCC 659
Qy 108 AlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThrValAsnAla 126
Db 660 TCAGCAACAATA-----TCCATCAGCGCAGCGCTCCTTCAGCGCATCTCAGCA 707

RESULT 3
B0804850
LOCUS      B0804850      709 bp      mRNA      linear      EST 31-JUL-2002
DEFINITION WHE3559_G01_M012S Wheat developing grains cdna library Triticum aestivum cdna clone WHE3559_G01_M01, mRNA sequence.
ACCESSION B0804850
VERSION   B0804850.1 GI:22029059
KEYWORDS  bread wheat.
SOURCE    Triticum aestivum
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; ; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 709)
AUTHORS   K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
TITLE     The structure and function of the expressed portion of the wheat genomes - Developing grains cdna library
JOURNAL   Unpublished (2002)
COMMENT   Contact: Olin Anderson
          US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
          800 Buchanan Street, Albany, CA 94710, USA
          Tel: 5105595773
          Fax: 5105595818
          Email: oandersn@nwp.usda.gov
          Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
          Seq primer: SK primer.
          Location/Qualifiers
FEATURES

```

source

1. 709

```

/organism="Triticum aestivum"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3559_G01_M01"
/clone_lib="wheat developing grains cdna library"
/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/note="Vector: Lambda ZAP II, excised phagemid; Site 1: EcoRI; Plants were grown under six following different environmental regimes in greenhouse, Environment 1) 240C/170C day/night, well-watered, with post-anthesis fertilizer, Environment 2) 240C/170C day/night, well-watered, without post-anthesis fertilizer, Environment 3) 370C/170C day/night, well-watered, post-anthesis fertilizer, Environment 4) 370C/170C day/night, well-watered, without post-anthesis fertilizer, Environment 5) 370C/170C day/night plus drought, with post-anthesis fertilizer, Environment 6) 370C/170C day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the TJ Clonase lab (Chin, Clonase, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."

```

```

BASE COUNT      234 a 215 c 113 g 147 t
ORIGIN

```

Alignment Scores:

```

Pred. No.:      0 632      Length:      709
Score:          94.50      Matches:      35
Percent Similarity: 37.96%      Conservative: 17
Best Local Similarity: 25.55%      Mismatches: 54
Query Match:    13.98%      Indels:      31
DB:             14      Gaps:      5

```

US-10-048-196-2 (1-134) x B0804850 (1-709)

```

Qy 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer 26
Db 139 GTATTCGCAACCAACAAGCAATTCACAGCAGCAACAACAACAACAACAACA 198
Qy 27 AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeLysHisAlaValLysGln 46
Db 199 AATCCTTCAACAATAATTTGCAACAACAACAACAACAACAACAACAACA 240
Qy 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTrpLysIle-----AlaAlaMet 63
Db 241 GGATGCTGCTGTGCACAAACCAACATAGCGCATCGCAAGTATTCGACACAAG 300
Qy 64 LysLeuSerSerGluSerLysAlaLysIleSer-----Lys 74
Db 301 TACTTACCAGCTATTGCAAGAATTTGTTGTCAGCACCTATGGCAGATCCCTGAGCAGTC 360
Qy 75 -----GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaVal 89
Db 361 GCAGTGCAGGCCCATCCAAATTTGTTGTCATGCTAT-----TATTCTGCA 405
Qy 90 SerLeuThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGln 109
Db 406 TCACACAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 465

```



```

/organism="Triticum aestivum"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone.lib="WHE3588_G02_M04"
/tissue_type="wheat developing grains cDNA library"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/notes="vector: Lambda ZAP II, excised phagemid; Site_1: EcoRI; Plants were grown under six following different environmental regimes in greenhouse, Environment 1) 24OC/17OC day/night, well-watered, with post-anthesis fertilizer, Environment 2) 24OC/17OC day/night, well-watered, without post-anthesis fertilizer, Environment 3) 37OC/17OC day/night, well-watered, with post-anthesis fertilizer, Environment 4) 37OC/17OC day/night, well-watered, without post-anthesis fertilizer, Environment 5) 37OC/17OC day/night plus drought, with post-anthesis fertilizer, Environment 6) 37OC/17OC day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pluescript SK(-) phagemids in the TJ Close lab (Chin, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."
BASE COUNT 260 a 252 c 119 g 153 t
ORIGIN
Alignment Scores:
Pred. No.: 0.754 Length: 784
Score: 94.50 Matches: 35
Percent Similarity: 37.96% Conservative: 17
Best Local Similarity: 25.55% Mismatches: 54
Query Match: 13.98% Indels: 31
DB: 14 Gaps: 5

US-10-048-196-2 (1-134) x BQ807264 (1-784)
QY 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer 26
Db 370 GTATTGCGCAACCAAGAACCAATTTACAGCAGCAGCAACCAACAACAACA 429
QY 27 AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLysHisAlaValLysGln 46
Db 430 AATCTCTCAACAATAATTTGCACACAACTGATTCATG-----CAT 471
QY 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIle-----AlaAlaMet 63
Db 472 GGATGTTGTTATTCACACACACACATAGCGCATGGAATCACAAGTTTCCACAAC 531
QY 64 LysLeuSerSerGluSerLysAlaLysIleSer----- 74
Db 532 TACTTACCAGCTTGCAGAGAATTGTTGTTCAGCACCTATGGCAGATCCCTGAGCAG 591
QY 75 -----GluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaVal 89
Db 592 GCAGTGCCAGGCCATCCAAATATGTTTCATCTCTAT-----TATTCTGCA 636
QY 90 SerLeuThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGln 109
Db 637 TCACAACAAAAAACAACAACAACCATCGAGCCAGGTCTCTCTTCCAAACAGCCCTCTGCA 696

```

```

QY 110 LysIleValArgHisSerLeuLysProCysMetLeuGluThrValAsnAla 126
Db 697 ACAATA-----TCCATTAGCGCAGGGCTCTTCCGGCATCTCAGCA 738

RESULT 6
BM802615 1080 bp mRNA linear EST 05-MAR-2002
LOCUS AGENCOURT_6459878 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:581015
5', mRNA sequence.
ACCESSION BM802615
VERSION BM802615.1 GI:19119438
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1080)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12340 row: d column: 08
High quality sequence stop: 740.
Location/Qualifiers
1..1080
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:581015"
/clone.lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 317 a 343 c 228 g 189 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 1.52 Length: 1080
Score: 94.00 Matches: 38
Percent Similarity: 44.17% Conservative: 15
Best Local Similarity: 31.67% Mismatches: 45
Query Match: 13.91% Indels: 22
DB: 14 Gaps: 4

US-10-048-196-2 (1-134) x BM802615 (1-1080)
QY 9 AlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer 27
Db 257 TCTACTGTCTCCAGCGCCACTAGTTTACCTTCTGTTGCCAGCACTCCCGCAGCTCCAAC 316
QY 28 ProLysAsnSerSerAlaAsn-LeuThrThrSerLeuLysHisAlaValLysGlnTh 47
Db 317 CCTGAAGCTCACCAGCAAACTGTGCTGTCCAGTCCCTCAAACTCCAAACCAAGCCAC 376
QY 47 rCysGlnThrGlnLeuThrGlyHis-GlnTyrTrpLysIleAlaAlaMetLysLeuSerS 67
Db 377 ACTCCAGGCCAAGCTC-----CACCTCGCCTGGAAATGGTTATCTCTCTGAATCCGGCA 430
QY 67 erGluSerLysAlaLysIleSerGluThrAlaCysGlyCys-----Vala 82
Db 431 GCAGTCACAGTCGCGGGTTCAGCGTCAGGCGCTGTGGCTGTGCCCCAGCTCTGACATGCT 490
QY 82 laAspLysAlaProGluAlaValSerLeuThrGluLeuThr----- 95

```

Db	491	CCAGCAGAACAGCTCAACACAGATGGCTGCACAGCAGCAACAAAGGGCCAAACTCATGCAG	550
Qy	96	-----ThralaAlaIleAsnProAsnAlaArgThrGluValAla	108
Db	551	CAGAAGCAGCAACAGCAACAGCAGCAGCAGCAGCAACAGCAGCAGCA	606
RESULT 7			
AG158466			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
PRIMERS			
Sequencing:			
LIBRARY			
Vector			
R.Site 1			
R.Site 2			
Location/Qualifiers			
1.			
/organism="Pan troglodytes"			
/db_xref="taxon:9598"			
/clone="RP43-023L05.TJ"			
/sex="male"			
/cell_type="lymphocytes"			
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"			
BASE COUNT			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-10-048-196-2 (1-134) x AG158466 (1-679)			
Qy	9	AlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer---	Asn 27
Db	162	TCTACTCTCCACGGCCACTAGTTTACTTCTTCTGCAGCACTCCCGCAGCTCCAAAC	221
Qy	28	ProlAsnSerSerAlaAsn-LeuThrThrSerLeuIleYshHisAlaValLysGlnTh	47
Db	222	CCTGCAGCTCAACAGCAACTGCTGTGTCAGTCCCTCAACTCCAAACCAAGCCAC	281
Qy	47	rCysGlnThrGlnLeuThrGlyHis-GlnTyrrTrpLysIleAlaAlaMetLysLeuSerS	67


```

444  Db      GATCCTTCACAAATTTTGCAACAACAACGATTCCATG-----CAG 485
      Qy      47  ThrCysGlnThrGlnLeuThrGlyHisGlnTyTrpLysIleAlaAlaMetLysLeuSer 66
      Db      486  GGAATGTTGATTGCAACAACAACAGCATAGCGCATGGGAAGCTCACAAAGTTT---GCAACA 542
      Qy      67  SerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaPro 86
      Db      543  AAGTACTTACCAGCTGGTGGCAACAATTTGTTGTGCACGACGCTGGGCAGAT-----CCC 596
      Qy      87  GluAlaValSerLeu----- 91
      Db      597  CGAGCAGCTCGGGTGCCAGCCATCCACATGTTGTTTCATGCTATTATTCTGCAATCAACA 656
      Qy      92  ---ThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLys 110

```

Db	657	ACAACAAACAACAACAACACCGTTCGAGCCAGTCTGCTCCACAGCTCTCAACAACA	716
Qy	111	IleValArgHisSerLeuLysProCysMetLeuGluThrValAsnAla	126
Db	717	ATA-----TCCATCAGCCAGGGCTCTTCCAGCCATCTCAGCA	755
RESULT 13			
LOCUS	BQ040628		
DEFINITION	BQ040628	812 bp mRNA linear	EST 31-JUL-2002
ACCESSION	WHE3556_H07_O14ZS	Wheat developing grains	CDNA library
VERSION	BQ040628	aestivum cDNA clone WHE3556_H07_O14,	maRNA sequence.
KEYWORDS	BQ040628.1	GI:22028791	
SOURCE	EST.		
ORGANISM	Bread wheat.		
REFERENCE	Triticum aestivum		
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.		
TITLE	1 (bases 1 to 812)		
JOURNAL	Altenbach, S., Anderson, O.D., Chao, S., Chin, A., Close, T.J., Cronin		
COMMENT	W.K., Crossman, C., Fenton, R.D., Lazo, G.R., Pham, J., Rausch, C.J., Wilson, C. and Woo, J.		
	The structure and function of the expressed portion of the wheat		
	genomes - Developing grains		
	CDNA library		
	Unpublished (2002)		
	Contact: Olin Anderson		
	US Department of Agriculture, Agriculture Research Service, Pacific		
	West Area, Western Regional Research Center		
	800 Buchanan Street, Albany, CA 94710, USA		
	Tel: 5105595773		
	Fax: 5105595818		
	Email: andersn@pw.usda.gov		
	Sequences have been trimmed		
	to remove vector sequence and low		
	quality sequence with phred score less than 20		
	Seq primer: SK primer.		
FEATURES	Location/Qualifiers		
source	1..812		
	/organism="Triticum aestivum"		
	/cultivar="Butte 86"		

```

/ub_xref= taxon:4300
/clone="WHE3556_H07_O14"
/clone_lib="Wheat developing grains cDNA library"
/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/note="Vector: Lambda ZAP II, excised phagemid; Site_1:
ECORI; Plants were grown under six following different
environmental regimes: in greenhouse, Environment 1)
24oC/17oC day/night, well-watered, with post-anthesis
fertilizer, Environment 2) 24oC/17oC day/night,
well-watered, without post-anthesis fertilizer,
Environment 3) 37oC/17oC day/night, well-watered, with
post-anthesis fertilizer, Environment 4) 37oC/17oC
day/night, well-watered, without post-anthesis fertilizer,
Environment 5) 37oC/17oC day/night plus drought, with
post-anthesis fertilizer, Environment 6) 37oC/17oC

```



```

Db 587 GGGCTTCTTCCAAACATCTCAGCA 610
      ||||| |||:|||||
BF429171 646 bp mRNA linear EST 29-NOV-2000
LOCUS WHE1709_C12_E23S wheat heat stressed spike cDNA library Triticum
DEFINITION aestivum cDNA clone WHE1709_C12_E23, mRNA sequence.
ACCESSION BF429171
VERSION BF429171.1 GI:11441095
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 646)
Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Malatrasi,M., Miller,R.,
Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
The structure and function of the expressed portion of the wheat
genomes - Heat stressed spike cDNA library
Unpublished (2000)
JOURNAL
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
FEATURES
Location/Qualifiers
source
1..646
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone_lib="WHE1709_C12_E23"
/tissue_type="wheat heat stressed spike cDNA library"
/dev_stage="Spikes at 5, 10, 15 and 20 days after
anthesis"
/lab_host="E. coli SOLR"
/Note="vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20
days after anthesis were heat stressed under two
conditions at Texas Tech University (D. Zhang in HT Nguyen
lab): (1) at 38 C for 4 hours and (2) 5 days of cyclic
treatment of 38 C for 4 hours. Total RNA and poly(A) RNA
were prepared, a cDNA library was made, and the cDNA
clones were in vivo excised to give phagescript phagemids
in the TJ Close lab (Choi, Close, Fenton, Malatrasi) at
the University of California, Riverside. plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
BASE COUNT 217 a 175 c 111 g 142 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 2,17 Length: 646
Score: 89.50 Matches: 36
Percent Similarity: 37.24% Conservative: 18
Best Local Similarity: 24.83% Mismatches: 52
Query Match: 13.24% Indels: 39
Dbs: 12 Gaps: 6
US-10-048-196-2 (1-134) x BF429171 (1-646)
Qy 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
||| ||||| ||| ||||| ||||| |||||
Db 2 GTATTCCGAACCAACACCAATTTCACGAGCAGCAGCAGCAACACACCA 61

```

```

Qy 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSer 38
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 ACAACAACAACAACAACAACAATCTTCAACAATAATTTTGAACAACAATGATTC 121
Qy 39 LeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrp 58
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 ATG-----CATGGATGTTGTTATTTGCAGCAACAACAATAGCGCATGG 163
Qy 59 LysIle-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer--- 74
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164 AGATCACAAAGTTTTCACAACAAGTACTTACCAGCTGTTGCAAGAATTTGTTGTCAGCA 223
Qy 75 -----GluThrAlaCysGlyCysVal 81
Db 224 CCTATGGCAGATCCCTGCAGCAGTCGTCAGGCCATCCAAATGTTGTTTCATGCTAT 283
Qy 82 AlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnPro 101
Db 284 -----TATTCTGCATCAACAACAACAACAACAACAACACCAACCATCGAGCCA 328
Qy 102 AsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeu 121
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 329 GGTGTCTCTTCCAAACAGCCTCTCGACAATA-----TCCATTAGCGCAGGCTCCTT 379
Qy 122 GluThrValAsnAla 126
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 CCGGCCATCTCAGCA 394
RESULT 16
BQ246193 BQ246193 671 bp mRNA linear EST 03-MAY-2002
LOCUS TAE15015A03R TAE15 Triticum aestivum cDNA clone TAE15015A03R, mRNA
sequence.
ACCESSION BQ246193
VERSION BQ246193.1 GI:20442069
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 671)
Cloutier,S.
Wheat functional genomics - Glenlea developing seeds cDNA libraries
Unpublished (2002)
JOURNAL Contact: Dr. Sylvie Cloutier
COMMENT Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 015 row: A column: 03
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
source
1..671
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone_lib="TAE15015A03R"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/Note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 15 days post-anthesis"
BASE COUNT 218 a 204 c 115 g 134 t
ORIGIN
Alignment Scores:

```



```

Db 640 TCAGCA 645
:::||||
RESULT 20
BQ246558
LOCUS
DEFINITION
BQ246558 694 bp mRNA linear EST 03-MAY-2002
TaE15008E10R TaE15 Triticum aestivum cDNA clone TaE15008E10R, mRNA
sequence.
ACCESSION
BQ246558
VERSION
BQ246558.1 GI:20442434
KEYWORDS
EST.
SOURCE
bread wheat.
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 694)
Cloutier,S.
Wheat functional genomics - Glenlea developing seeds cDNA libraries
Unpublished (2002)
Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 008 row: E column: 10
Seq primer: M13 Reverse.
Location/Qualifiers
1..694
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaE15008E10R"
/clone_lib="TaE15"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/notes="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 15 days post-anthesis"
BASE COUNT 232 a 227 c 100 g 135 t
ORIGIN
Alignment Scores
Pred. No.: 2.84 Length: 694
Score: 89.00 Matches: 36
Percent Similarity: 38.03% Conservative: 18
Best Local Similarity: 25.35% Mismatches: 52
Query Match: 13.17% Indels: 36
DB: 14 Gaps: 6
US-10-048-196-2 (1-134) x BQ246558 (1-694)
QY 7 ValThraAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
||| ||||| ||| ||||| ||||| |||||
Db 159 GTATTCCGACACCAACCAACCAATTTCCAGCAGCAGCAGCAGCAGCAACACACA 218
QY 22 SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLys 41
::: ||| ||||| ||| ||| ||| ||| |||
Db 219 ACAACAACAACAACAATCTTCAACAATTTGCCAACAACAATGATTCATG----- 272
QY 42 HisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIle--- 60
||| ||| ||| ||| ||| ||| ||| |||
Db 273 -----CATGGATGTTGATTGCAGCAACAACATAGCCGATGGAAGATCACA 320
QY 61 -----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer----- 74
||| ||| ||| ||| ||| ||| ||| |||
Db 321 AGTTTGCACAACAAAGTACTTACCAGCTGTTGCAAGATTGTTGTCAGCAGCACCTATGGCA 380
::: ||| ||| ||| ||| ||| ||| ||| |||
GlutThraAlaCysGlyCysValAlaAspLys 84
::: ||| ||| ||| ||| ||| ||| ||| |||
Db 381 GATCCCTGACAGTCGCGAGTCCAGGCCATCCCAAGTTGTTTCATGCTAT----- 431
QY 85 AlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArg 104
::: ||| ||| ||| ||| ||| ||| ||| |||
Db 432 -----TATTCTGCTCAACAACAACAAACAAACCAACCAACCAACCAACCAACCACTCCTT 485
QY 105 ThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThrVal 124
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 486 CCAACAGCCTCTGCAACAATA-----TCCATTAGGCCAGGGCTCCTTCGCGGCATC 536
QY 125 AsnAla 126
::: ||| ||| ||| ||| ||| ||| ||| |||
Db 537 TCAGCA 542
RESULT 21
BQ246454
LOCUS
DEFINITION
BQ246454 731 bp mRNA linear EST 03-MAY-2002
TaE15010E09R TaE15 Triticum aestivum cDNA clone TaE15010E09R, mRNA
sequence.
ACCESSION
BQ246454
VERSION
BQ246454.1 GI:20442330
KEYWORDS
EST.
SOURCE
bread wheat.
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 731)
Cloutier,S.
Wheat functional genomics - Glenlea developing seeds cDNA libraries
Unpublished (2002)
Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 010 row: E column: 09
Seq primer: M13 Reverse.
Location/Qualifiers
1..731
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaE15010E09R"
/clone_lib="TaE15"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/notes="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 15 days post-anthesis"
BASE COUNT 251 a 235 c 111 g 134 t
ORIGIN
Alignment Scores
Pred. No.: 3.11 Length: 731
Score: 89.00 Matches: 36
Percent Similarity: 38.03% Conservative: 18
Best Local Similarity: 25.35% Mismatches: 52
Query Match: 13.17% Indels: 36
DB: 14 Gaps: 6
US-10-048-196-2 (1-134) x BQ246454 (1-731)
QY 7 ValThraAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
||| ||||| ||| ||||| ||||| |||||
Db 261 GTATTCCGACACCAACAACCAACCAATTTCCAGCAGCAGCAGCAGCAACCAACA 320

```



```

QY 22 SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLeuLys 41
Db 321 ACAACAACAACAACAATCCTTCAACAATTTTGCACAACAACAATGATCCCATG----- 374
QY 42 HisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIle----- 60
Db 375 -----CATGGATGTTGTTATTCGAGCAACAACAATAGGCATGGAAGATCACA 422
QY 61 -----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer----- 74
Db 423 AGTTTTCACAACAAGTACTTACCAGCTGTTGCAAGAAATGTTGTCAGCACCTATGGCA 482
QY 75 -----TATTCGTCATCAACAACAACAACAACAACAACAACCATCGAGCGCTCCCTT 587
Db 483 GATCCCTGAGCAGTCGAGTCGAGGCCATCCACAAGTGTTCATGCTAT----- 533
QY 85 AlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArg 104
Db 534 -----TATTCGTCATCAACAACAACAACAACAACAACAACCATCGAGCGCTCCCTT 587
QY 105 ThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThrVal 124
Db 588 CCAAGAGCCTTCGACACATA-----TCCATGAGCCAGGCTCCTTCGCGCCATC 638
QY 125 AsnAla 126
Db 639 TCAGCA 644

RESULT 22
LOCUS BQ839009
DEFINITION WHE3590_C09_F18z2s Wheat developing grains cDNA library Triticum aestivum CDNA clone WHE3590_C09_F18, mRNA sequence.
ACCESSION BQ839009
VERSION BQ839009.1
KEYWORDS 792 bp mRNA linear EST 08-AUG-2002
SOURCE Triticum aestivum
ORGANISM Triticum aestivum
REFERENCE 1 (bases 1 to 792)
AUTHORS Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronin,K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
TITLE The structure and function of the expressed portion of the wheat genomes - Developing grains cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@wv.usda.gov
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.

FEATURES
source
1..792
    /organism="Triticum aestivum"
    /cultivar="Butte 86"
    /db_xref="taxon:4565"
    /clone="WHE3590_C09_F18"
    /clone_lib="Wheat developing grains cDNA library"
    /tissue_type="whole grains"
    /dev_stage="3-44 days post anthesis seed"
    /lab_host="E. coli SOLR"
    /note="Vector: Lambda ZAP II, excised phagemid; Site_1: EcoRI; Plants were grown under six following different environmental regimes in greenhouse, Environment 1) 240C/170C day/night, well-watered, with post-anthesis

```

```

fertilizer, Environment 2) 240C/170C day/night, well-watered, without post-anthesis fertilizer, Environment 3) 370C/170C day/night, well-watered, with post-anthesis fertilizer, Environment 4) 370C/170C day/night, well-watered, without post-anthesis fertilizer, Environment 5) 370C/170C day/night plus drought, with post-anthesis fertilizer, Environment 6) 370C/170C day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the TJ Close lab (Chin, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others).
BASE COUNT 266 a 257 c 123 g 146 t
ORIGIN

Alignment Scores:
Pred. No.: 4,12 Length: 792
Score: 88.50 Matches: 36
Percent Similarity: 37.84% Conservatives: 20
Best Local Similarity: 24.32% Mismatches: 51
Query Match: 13.09% Indels: 41
DB: 14 Gaps: 7

US-10-048-196-2 (1-134) x BQ839009 (1-792)
QY 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla-----Ser 22
Db 381 GTATTGCGAACCAACAACCAATTTTCGAGCAGCAGCAGCAGCAGCAGCAACAACA 440
QY 23 ThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLysHis 42
Db 441 ACAACAACAACAATAATCTTACAACAATTTTGCACAACAACAATGATTCATG----- 491
QY 43 AlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIle----- 60
Db 492 -----CAGGGATGTTGTTATTGCAACAACAACAATAGGCATGGAAGTCAAGT 542
QY 61 -----AlaAlaMetLysLeuSerSerGluSerLysAla 71
Db 543 TTTGCAAGAAAGTACTTACCAGCTGTCGCAACAATTTGTGTGTC-----GCA 590
QY 72 LysIleSerGlu-----ThrAlaCysGlyCys--- 80
Db 591 GCTGTGGCAGATCCCGCAGCAGTCGCGGTGCCAAGCATCCCAATGTTGTCGTAT 650
QY 81 -----ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrAlaAla 98
Db 651 TATTCTGCATCAACAACACCACCACCACCACAACAACAACAACAACAACAACC 710
QY 99 IleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysPro 118
Db 711 GTTGAGCAAGTCTCCTTCAACAGCCTCAGCAACAATA-----TCCATCAGGCCA 761
QY 119 CysMetLeuGluThrValAsnAla 126
Db 762 GGGCTTCTTTTCAACATCTCAGCA 785

RESULT 23
LOCUS BQ262346
DEFINITION WHE0874_G03_N06Zs Wheat 20-45 DAP spike cDNA library Triticum

```


Db	177	GTATTCCGACCAACCAACCAATTTTCACAGCAGCAGCAGCAGCAGCAGCAACA	236
Qy	22	-----SerThrProGluSerAsnProLysSerSerAlaAsnLeuThrThrSerLeu	39
Db	237	ACACACACACACACACACAAATTCCTTCAACAATTTTGCACACACAACTGATTCATG	296
Qy	40	IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLys	59
Db	297	-----CATGGATGTTGTTATTCGACGACACACACATAGCGCATGGAAG	338
Qy	60	Ile-----AlaAlaMetLysSerSerGluSerLysAlaLysIleSer-----	74
Db	339	ATCAAGATTGTTGCAACAAAGTACTTACCAGCTGTTGCAAGATTTGTTGTCACACACCT	398
Qy	75	-----GluThrAlaCysGlyCysValala	82
Db	399	ATGCGAGATCCTGAGCAGTCGCGAGTGCCAGGCGCATCCACATGTTGTTCATGTAT---	455
Qy	83	AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn	102
Db	456	-----TATTCTGCATCAACACACAAAAACACACACACACATCGACCCAGGT	503
Qy	103	AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu	122
Db	504	CTCCTTGCAACAGCGCTCTGCAACAATA-----TCCATTAGCGCAGGCGCTCTTCG	554
Qy	123	ThrValAsnAla	126
Db	555	GCCATCTCAGCA	566
RESULT	25		
LOCUS	BM634463		
DEFINITION	1700687509132 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone	714 bp	mRNA linear EST 26-FEB-2002
ACCESSION	19600449653777	5', mRNA sequence.	
VERSION	BM634463		
KEYWORDS	BM634463.1	GI:18933974	
SOURCE	EST.		
ORGANISM	African malaria mosquito.		
REFERENCE	Anopheles gambiae		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.		
TITLE	1 (bases 1 to 714)		
JOURNAL	Holt R.A., Lin J.-J., Murphy S.D., Evans C.A., Kraft C.L., Charlab R., Collins F.H., Venter J.C. and Hoffman S.L.		
COMMENT	Celera Anopheles gambiae EST project Unpublished (2002) Contact: Holt R.A. Celera Genomics 45 W. Gude Dr., Rockville, MD 20850, USA Tel: 240533151 Fax: 240534580 Email: HoltRA@celera.com Plate: NU01004A9N row: C column: 23 Seq primer: M13 Reverse.		
FEATURES	Location/Qualifiers		
source	1..714		
	/organism="Anopheles gambiae"		
	/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"		
	/db_xref="taxon:7165"		
	/clone="19600449653777"		
	/clone_lib="A.Gam.ad.cDNA1"		
	/dev_stage="Adult"		
	/lab_host="DH10b"		
	/note="vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"		

```

/organism="Triticum aestivum"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3591.C06.E11"
/clone_lib="wheat developing grains cDNA library"
/tissue_type="whole grains"
/dev_stage="3-4 days post anthesis seed"
/lab_host="E. coli SOLR"

notes="Vector: Lambda ZAP II, excised phagemid; Site 1:
EcoRI; Plants were grown under six following different
environmental regimes in greenhouse, Environment 1)
240C/170C day/night, well-watered, with post-anthesis
fertilizer, Environment 2) 240C/170C day/night,
well-watered, without post-anthesis fertilizer,
Environment 3) 370C/170C day/night, well-watered, with
post-anthesis fertilizer, Environment 4) 370C/170C
day/night, well-watered, without post-anthesis fertilizer,
Environment 5) 370C/170C day/night plus drought, with
post-anthesis fertilizer, Environment 6) 370C/170C
day/night plus drought, without post-anthesis fertilizer,
developing wheat grains from the following were excised
and frozen in liquid nitrogen, Environment 1 at 3, 5, 7,
8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment
2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44
DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28,
32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20,
24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16,
20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16
20, 24, 28, 30 DPA and total RNA was prepared by S.
Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA
library was made using poly (A) RNA, and the cDNA clones
were in vivo excised to give pBluescript SK(-) phagemids
in the TJ Close lab (Chin, Close, Fenton) at the
University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (others)."
BASE COUNT      263 a 269 c 131 g 145 t
ORIGIN
Alignment Scores:
Pred. No.:      5.64      Length:      808
Score:          87.50      Matches:      36
Percent Similarity: 37.84%      Conservative: 20
Best Local Similarity: 24.32%      Mismatches: 51
Query Match:     12.94%      Indels:      41
DB:              14        Gaps:       7

US-10-048-196-2 (1-134) x BQ838664 (1-808)
QY 7 ValThraLathrLeuMetThrAlaPheThrLeuAlaSerCysAla-----Ser 22
Db 317 GTATTGCGAACCAACCAACCAATTTCGCAGCAGCAGCAGCAGCAGCAACAACCA 376
QY 23 ThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLysHis 42
Db 377 ACAACACACAACTCTACACAAATTTTGCACAACTGATCCATG-----427

QY 43 AlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIle-----60
Db 428 -----CAGGGATGTTGATTGCAACACACACACATAGCGCATGGAGCTCACAAAGT 478
QY 61 -----AlaAlaMetLysLeuSerSerGluSerLysAla 71
Db 479 TTTGCAAGAAAGTACTTACCAGCTGGTGCAACAATGTGTGTGCA-----GCA 526

QY 72 LysIleSerGlu-----ThrAlaCysGlyCys---80
Db 527 GCTGTGGCAGATCCCGAGCAGCTGGCGGTGCCAAGCATCCCAATGTGTGTCATGCTAT 586
QY 81 -----ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrAlaAla 98
Db 587 TATTCGTCATCACACACACACACACACACACACACACACACACACACACACAC 646

```

```

QY 99 IleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysPro 118
Db 647 GTTGAGCAAGTCTCTTCCACAGCCTCAGCAACATA-----TCCATCAGGCCA 697

QY 119 CysMetLeuGluThrValAsnAla 126
Db 698 GGGCTTCTTCCACACCATCTCAGCA 721

RESULT 27
LOCUS      BJ237941/c      849 bp      mRNA      linear      EST 05-APR-2002
DEFINITION      BJ237941 Y. Ogihara unpublished cDNA library, WHe Triticum
aestivum cDNA clone whello22 3', mRNA sequence.
ACCESSION      BJ237941
VERSION        BJ237941.1 GI:20053246
KEYWORDS       EST.
SOURCE         Bread wheat.
ORGANISM       Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.
REFERENCE      1 (bases 1 to 849)
AUTHORS        Ogihara Y. and Murai K.
TITLE          Expressed genes in Triticum aestivum
JOURNAL        Unpublished (2002)
COMMENT        Contact: Tadasu Shin-i.
                Center For Genetic Resource Information
                National Institute of Genetics
                1111 Yata, Mishima, Shizuoka 411-8540, Japan
                Tel: 81-559-81-6856
                Fax: 81-559-81-6855
                Email: tshin@genes.nig.ac.jp.
                Location/Qualifiers
FEATURES
source
    1..849
     /organism="Triticum aestivum"
     /cultivar="Chinese Spring"
     /db_xref="taxon:4565"
     /clone_lib="Y. Ogihara unpublished cDNA library, WHe"
     /clone="whello22"
     /tissue_type="seed DPA10"
     /dev_stage="Feekes' scale 11.2"
     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
     Site_1: EcoRI; Site_2: XhoI; Plants were grown under
     hydroponic conditions at UC Davis, salt stressed for 12
     hours, and for 7 days, then dissected and frozen (Akhunov
     in J Dvorak Lab). Total RNA was prepared from sheath
     tissue, equal quantities of RNA were pooled from the two
     samples, polyA was purified from the pooled RNA, a cDNA
     library was made, and the cDNA clones were in vivo
     excised to give pBluescript phagemids in the TJ Close lab
     at the University of California, Riverside (Akhunov, Chin
     , Choi, Close, Fenton, Kianian, Otto, Simons, Zhang).
     Plasmid DNA preparations and DNA sequencing were
     performed in the OD Anderson lab (all other authors)."
```

```

BASE COUNT      177 a 140 c 246 g 281 t 5 others
ORIGIN
Alignment Scores:
Pred. No.:      6.16      Length:      849
Score:          87.50      Matches:      39
Percent Similarity: 38.41%      Conservative: 14
Best Local Similarity: 28.26%      Mismatches: 50
Query Match:     12.94%      Indels:      35
DB:              13        Gaps:       7

```

US-10-048-196-2 (1-134) x BJ237941 (1-849)

```

QY 7 ValThraLathrLeuMetThrAlaPheThrLeuAlaSerCysAla-----Ser 22
Db 685 GTATTGCGAACCAACCAACCAATTTCAGCAGCAGCAGCAGCAGCAACAACCA 626
QY 23 ThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLysHis 42

```

```

Db 625 ACACAAACACAGATCTCTCAACAATTTTGCACAATTAAGTATTCATG----- 575
QY 43 AlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAla---- 61
Db 574 -----CAGGATGTTGTATTGCAACAACACAGCATAGCGCATGGAGCTCACAAGT 524
QY 62 ---AlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCys 80
Db 523 TTGCAACAAGTACTTACCAGCTGGTGCACAATTTGTGTGCAACAGCT----- 473
QY 81 ValAlaAspLysAlaProGluAlaVal----- 89
Db 472 GTGGCAGATCCCGGAGCATCGCGGTGCACAGCCATCCCAATGTTGTTTCATGCTATTAT 413
QY 90 -----SerLeuThrGluLeuThrAlaAlaIleAsn-----ProAsnAla 103
Db 412 TCTGCATCAACAACAACAACAACAACACCGTTGAGCCAGGTCTGCTTCCACAGCC 353
QY 104 ArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeu 121
Db 352 TCAACAACAATATCC-----ATCAGGCGAGGCTCTTCCAGCCATCTTA 308

RESULT 28
LOCUS BE423599 536 bp mRNA linear EST 24-JUL-2000
DEFINITION WHE0072_H10_020S Wheat endosperm cDNA library Triticum aestivum
cDNA clone WHE0072_H10_020, mRNA sequence.
ACCESSION BE423599
VERSION BE423599.1 GI:9421442
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 536)
AUTHORS Altenbach,S., Anderson,O.D., Chao,S., Galili,G., Han,P.S., Hsia
,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and
Tong,J.C.
The structure and function of the expressed portion of the wheat
genomes - Endosperm cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
FEATURES
Location/Qualifiers
source
1..536
/organism="Triticum aestivum"
/cultivar="Cheyenne"
/db_xref="taxon:4565"
/clone_lib="Wheat endosperm cDNA library"
/tissue_type="Endosperm"
/dev_stage="5 to 30 days post anthesis seed"
/lab_host="E. coli SOLR"
/note="Vector: Lambda ZAP II, excised phagemid; Site_1:
EcoRI; Seeds collected, endosperm isolated, and RNA
prepared by Susan Altenbach. Library constructed by
Stratagene, Inc. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab."
BASE COUNT 186 a 173 c 75 g 102 t
ORIGIN
Alignment Scores: 3.15 Length: 536
Pred. No.: 87.00 Matches: 36
Score:

```

```

Percent Similarity: 36.81% Conservative: 17
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 12.87% Indels: 38
DB: 10 Gaps: 6

US-10-048-196-2 (1-134) x BE423599 (1-536)

QY 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
Db 139 GTATTGCGAACCAACAACCAATTTTCAGCAGCAGCAGCAGCAGCAGCAACA 198
QY 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
Db 199 ACAACAACAACAACAACAACAATCTTCAACAATTTTGCACAACAACACTGATTCATG 258
QY 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLys 59
Db 259 -----CATGGATGTTGTATTGTCAGCAGCAGCAGCAGCAGCATGGAG 300
QY 60 Ile-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer----- 74
Db 301 ATCACAAGTTTGTCAACAAGTACTTACCAGCTGTTTGCAGAATTTGTGTTCAACACCT 360
QY 75 -----GluThrAlaCysGlyCysValAla 82
Db 361 ATGGCAGATCCCTGAGCAGTGCAGTGCAGGCCATCCCAATGTTGTTTCATGCTAT--- 417
QY 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102
Db 418 -----TATTCTGCATCAACAACAACAACAACAACAACACCATCGAGCCAGGT 465
QY 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
Db 466 CTCCTTCCCAACAGCTCTGCAACAATA-----TCCATTAGGCCAGGCTCCTTCG 516
QY 123 ThrValAsnAla 126
Db 517 GCCATCTCAGCA 528

RESULT 29
LOCUS BQ246932 610 bp mRNA linear EST 03-MAY-2002
DEFINITION TAE15003D12R TAE15 Triticum aestivum cDNA clone TAE15003D12R, mRNA
sequence.
ACCESSION BQ246932
VERSION BQ246932.1 GI:20442808
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 610)
AUTHORS Cloutier,S
TITLE Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 003 row: D column: 12
Seq primer: M13 Reverse.
Location/Qualifiers
source
1..610
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TAE15003D12R"

```

```

/clone_lib="Tag15"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/note="vector: pCMV-SPORTS.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 15 days post-anthesis"
203 a 201 c 90 g 116 t

```

Alignment Scores:	
Pred. No.:	3 96
Score:	87.00
Percent Similarity:	36.81%
Best Local Similarity:	25.00%
Query Match:	12.87%
DB:	14
Length:	610
Matches:	36
Conservative:	17
Mismatches:	53
Indels:	38
Gaps:	6

US-10-048-196-2 (1-134) x BQ246932 (1-610)

Oy 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
 ||| ||||| ||| ||||| |||::: |||
Db 172 GTATTTCGCACACCACAAACAATTCACAGCAGCAGCAGCAGCAGCAACA 231

Qy 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
 :::| | | :| | | | | :| | | | | :| | | | |
Db 232 ACAACAACCAACAACAACAATCTTCAACAATTTTGCAACAACAACACTGATTCCATG 291

Qy 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLys 59
Dδ 292 -----CATGGATGTTGATTGCAGCACACAAACATAGCGCATGGAG 333

Qy 60 Ile-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer----- 74
||| ||| ||| :: : :::: : ::|||
Db 334 ATCACAGATTTCACAAAGAAGTACTACCAGCTGTGGCAGGAATTGCTGTGCACACACCT 393

Qy 75 -----GluThrAlaCysGlyCysValAla 82
 |||
Db 394 ATGGCAGATCCCTGAGCAGTCGCCAGTGCACGCCCATCCACATGTTCATGCTAT--- 450

Qy 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102
::: ||| |||::: ||||| ||| |||
Db 451 -----TATTTCGTCATCAACAACAAAAAACAAACAACCAACCATCGAGCCAGST 498

QY 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
::: ||| |||::: |||::: |||
Db 499 CTCCTTCCACACGCTCTGCAACATA-----TCCATTAGGCCAGGGCTCTCCG 549

Qy 123 ThrValAsnAla 126
 :::>:::
Db 550 GCCATCTCAGCA 561

RESULT 30					
BQ244975					
LOCUS	631 bp	mRNA	linear	EST 03-MAY-2002	
DEFINITION	BQ244975	Triticum aestivum cDNA clone TaeI5032G05R,			
					sequence.

ACCESSION	BQ244975	
VERSION	BQ244975.1	GI:20440851
KEYWORDS	EST.	
SOURCE	bread wheat.	
ORGANISM	Triticum aestivum	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae	
	; Triticaceae; Triticum.	

REFERENCE	1 (bases 1 to 631)
AUTHORS	Cloutier, S.
TITLE	Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL	Unpublished (2002)
COMMENT	Contact: Dr. Sylvie Cloutier Cereal Research Centre, Agriculture and Agri-food Canada 195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9 Tel: (204) 983-2340

Tel: (204) 983-2340

Fax: (204) 983-4504
Email: scioutier@agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 032 row: G column: 05
Seq primer: M13 Reverse.

FEATURES

```

/organism="Triticum aestivum"
/cultivar="Glenles"
/db_xref="taxon:4565"
/clone="raE15032G05R"
/clone_lib="raE15"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"

```

[illegible]

Alignment Scores:	
Pred. No.:	4.2
Score:	87.00
Percent Similarity:	37.67%
Best Local Similarity:	26.03%
Query Match:	12.87%
DB:	14
	7
	Gaps:
	Indels:
	Mismatches:
	Conservative:
	Matches:
	Length:

US-10-048-196-2 (1-134) x BQ244975 (1-631)

Qy	7	ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrPro-----	24
Db	98	GTATTGCAACCCACACAAACCAATTTACAGCAGCAGCAGCAGCAGCAACAACAACA	157

Qy 25 -----GluserAshnProLysAsnSerSerAlaAsnLeuThrThr 37
 :::||||| ::: ||| |||
Db 158 ACAACAACACACACACAGAACAAATCTTCAACAATAATTGCAACAACAACACTGAC 217

38 SerLeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
 |||::: ||| ||| ||| |||
 218 TCCATG-----CATGGATGTTGTATTGCAAGCAACACAAACATAGCGCG 259

Qy 58 TrpLysIle-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer 74
||||| ||| ||| :: :|||
Db 260 TGGAAAGTCACAAAGTTTGC AACAAAGTACTTACCAGCTGTGTGCAAGAATTTGTGTGTGCA 319

Qy 75 -----GluThrAla-----CysGlyCys 80
 ||| ||| ||| |||
Db 320 GCACCATGGCACATCCCTGAGAAAGTTGCAGTGGCAGGCCATCCACAATGTGTTCATGC 379

QY 81 ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
::: ||| |||::: |||||
Db 380 TAT-----TATTGTCATCAACACAAAAACACACACACATATCGAG 424

OY 101- proAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMet 120
 ||| , ::|| ||| |::| :|| :||
D6 425 CCACCGCTCCCTCCACACACCGCTCCACACACATA-----TCCATTTACCCCGGGTC 475

Qy 121 LeuGluThrValAsnAla 126
||| ::::||||
db 476 GTTCCCCCCCCTATGGCGCCTA 493

RESULT 31
AW448815

DEFINITION	BRY_1500 BRY <i>Triticum aestivum</i> cDNA clone P52-1L, mRNA sequence
ACCESSION	AW448815
VERSION	AW448815.1 GI:12019350

SOURCE bread wheat.

```

ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 641)
AUTHORS     Clarke,B.C., Hobbs,M. and Appels,R.
TITLE        Genes active in developing wheat endosperm
JOURNAL      Unpublished (2000)
COMMENT      Contact: Bryan Clarke
              Division of Plant Industry
              C.S.I.R.O.
              GPO Box 1600, Canberra, ACT, Australia
              Tel: 61 2 6246 5054
              Fax: 61 2 6246 5000
              Email: bryan@dpi.csiro.au.

FEATURES   Location/Qualifiers
            source
              1..641
              /organism="Triticum aestivum"
              /cultivar="Wynona"
              /db_xref="taxon:4565"
              /clone_lib="P52-IL"
              /clone_lib="BRY"
              /cell_type="endosperm"

BASE COUNT 215 a 202 c 101 g 123 t
ORIGIN
Alignment Scores:
Pred. No.: 4.32 Length: 641
Score: 87.00 Matches: 38
Percent Similarity: 37.67% Conservativeness: 17
Best Local Similarity: 26.03% Mismatches: 51
Query Match: 12.87% Indels: 40
DB: 10 Gaps: 7

US-10-048-196-2 (1-134) x AW448815 (1-641)
Qy 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrPro----- 24
Db 107 GTATTCGAACCAACAAACAAATTTTCACAGCAGCAGCAGCAGCAGCAACAAACAA 166
Qy 25 -----GlusAsnProLysAsnSerSerAlaAsnLeuThrThr 37
Db 167 ACAACAACAACAACAAGAACAAATCCTTCAACAATTTTGCACAAACAACATGAC 226
Qy 38 SerLeuLeLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
Db 227 TCCATG-----CATGGATGTTGTATTGCGACCAACAACATAGCGCG 268
Qy 58 TrpLysIle-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer 74
Db 269 TGGAGATCAACAAGTTTTCACAAAGTACTTACCAGCTGTTGCAAGATTTGCTGTCA 328
Qy 75 -----GluThrAla-----CysGlyCys 80
Db 329 GCACCTATGCGAGATCCCTGAGAGATTCAGTCGCGCCGATCCACATGTTGTTATGCG 388
Qy 81 ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrAlaAlaIleasn 100
Db 389 TAT-----TATTCTGCATCAACAACAACAAACAACAACAACATATCGAG 433
Qy 101 ProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMet 120
Db 434 CCAGGTCTCTCCCTCCAAACAGCCTCAGCAACAATA-----TCCATTAGGCCAGGGCTC 484
Qy 121 LeuGluThrValAsnAla 126
Db 485 CTTCGGCCATCTCAGCA 502

RESULT 32
BI968881
LOCUS      BI968881
DEFINITION GM830006B11H03 Gm-r1083 Glycine max cDNA clone Gm-r1083-2285 3',
            mRNA sequence.

```

```

ACCESSION  BI968881
VERSION     BI968881.1
KEYWORDS    GI:16343286
SOURCE      soybean.
            Glycine max
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1 (bases 1 to 641)
AUTHORS     Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Expeliding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
TITLE        A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL      Unpublished (1999)
COMMENT      Other_ESTs: AW203461 corresponding to Gm-cl028-1497 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Incyte Genomics, 4633 World
            Parkway Circle St. Louis, Missouri 63134. Phone (800) 430-0030 or
            (314) 427-3222 FAX: (314) 427-3324. Web site:
            http://www.incyte.com/reagents/catalog.jsp?page=clones/collaboratio
            n/index
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
            Location/Qualifiers
            1..641
            /organism="Glycine max"
            /db_xref="taxon:3847"
            /clone="Gm-r1083-2285"
            /clone_lib="Gm-r1083"
            /note="The library Gm-r1083 is a sequence-driven, reracked
            set of 4,992 clones selected from cDNA libraries from
            various tissues and stages of development of soybean. It
            represents 1117 sequences from the progenitor library
            Gm-cl009 (from mature roots of 2 month old greenhouse
            grown 'Williams' soybean plants); 820 sequences from the
            progenitor library Gm-cl013 (from 2 to 3 week old whole
            plants of Williams); and 3055 sequences from library
            Gm-cl028 (from 'Supernod' plants whose seedlings were
            inoculated with Bradyrhizobium japonicum, courtesy of Dr.
            Gary Stacey). The 5' ESTs of the source clones from the
            different progenitor libraries was used to select
            singletons, or a representative of each contig, which were
            reracked to form library Gm-r1083. The cDNA clones of the
            reracked Gm-r1083 library were then sequenced at the 3'
            end. The contig analysis to select unique genes was
            performed by the laboratory of Ernest Retzel, Center for
            Computational Genomics and Bioinformatics, University of
            Minnesota, http://web.ahc.umn.edu/biodata/nsfsoy/.
            Reracking was performed by Incyte Genomics, St. Louis,
            http://www.incyte.com, and 3' sequencing by the Keck
            Center for Comparative and Functional Genomics, University
            of Illinois, http://www.lie.uiuc.edu/biotech/keck.html.
            Note: The corresponding 5' EST from each clone in the
            Gm-r1083 library is listed in the 'OTHER EST' field. The
            detailed information on the source library for each clone
            can also be obtained by referring to the Incyte Genomics
            clone ID of the original cDNA library that is also listed
            under 'OTHER EST'."

BASE COUNT 194 a 155 c 123 g 145 t 24 others
ORIGIN
Alignment Scores:
Pred. No.: 4.32 Length: 641
Score: 87.00 Matches: 40
Percent Similarity: 40.88% Conservativeness: 16
Best Local Similarity: 29.20% Mismatches: 55

```


/clone="TaeE25023A08R"
/clone_lib="TaeE25"
/tissue_type="developing_seeds"
/dev_stage="25 days after anthesis"
/lab_host="E. coli DH10B"
/note="vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 25 days post-anthesis"

BASE COUNT 206 a 204 c 106 g 127 t
ORIGIN

Alignment Scores:
Pred. No.: 4.34 Length: 643
Score: 87.00 Matches: 36
Percent Similarity: 36.81% Conservative: 17
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 12.87% Indels: 38
DB: 14 Gaps: 6

US-10-048-196-2 (1-134) x BQ251877 (1-643)

Qy 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
Db 101 GTATTGCGAACCAACAAACCAATTTACAGCAGCAGCAGCAGCAGCAGCAACA 160
Qy 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
Db 161 ACAACAACAACAACAACAACAATCTTCAACAAATTTGCAACAACAACACTGATTCATG 220
Qy 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLys 59
Db 221 -----CATGGATGTTGATTGCGAGCAACAACAACATAGCGCATGGGAAG 262
Qy 60 Ile-----AlaAlaMetLysLeuSerGluSerLysAlaLysIleSer----- 74
Db 263 ATCACAAGTTTGGCAACAAGTACTTACCAGCTGTTGCAAGAATGTGTGTCAACACCT 322
Qy 75 -----GluThrAlaCysGlyCysValAla 82
Db 323 ATGGCAGATCCCTGAGCAGTCGAGTGCCAGGCCATCCACAATGTTGTCATGCTAT--- 379
Qy 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102
Db 380 -----TATTCGTCATCAACAACAACAACAACAACAACATCAGCGCCAGGT 427
Qy 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
Db 428 CTCCTTCCAACAGCCTCTGCAACAATA-----TCCATTAGGCCAGGGCTCCTCCG 478
Qy 123 ThrValAsnAla 126
Db 479 GCCATCTCAGCA 490

RESULT 35
BQ246079 BQ246079 647 bp mRNA linear EST 03-MAY-2002
LOCUS TaeI5016E10R TaeI5 Triticum aestivum cDNA clone TaeI5016E10R, mRNA
DEFINITION sequence.

ACCESSION BQ246079
VERSION BQ246079.1 GI:20441942
KEYWORDS EST.
SOURCE bread wheat.

ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 647)
AUTHORS Cloutier, S.
TITLE Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL Unpublished (2002)
CONTACT: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2N9

Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 016 row: E column: 10
Seq primer: M13 Reverse.
Location/Qualifiers
1. 647

FEATURES
source

/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone_lib="TaeI5016E10R"
/clone_lib="TaeI5"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/note="vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 15 days post-anthesis"

BASE COUNT 217 a 197 c 105 g 128 t
ORIGIN

Alignment Scores:
Pred. No.: 4.39 Length: 647
Score: 87.00 Matches: 36
Percent Similarity: 36.81% Conservative: 17
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 12.87% Indels: 38
DB: 14 Gaps: 6

US-10-048-196-2 (1-134) x BQ246079 (1-647)

Qy 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
Db 67 GTATTGCGAACCAACCAATTTTACAGCAGCAGCAGCAGCAGCAGCAACA 126
Qy 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
Db 127 ACAACAACAACAACAACAATCTTCAACAAATTTGCAACAACAACATGATTCATG 186
Qy 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLys 59
Db 187 -----CATGGATGTTGATTGCGAGCAACAACAACATAGCGCATGGGAAG 228
Qy 60 Ile-----AlaAlaMetLysLeuSerGluSerLysAlaLysIleSer----- 74
Db 229 ATCACAAGTTTGGCAACAAGTACTTACCAGCTGTTGCAAGAATGTGTGTCAACACCT 288
Qy 75 -----GluThrAlaCysGlyCysValAla 82
Db 289 ATGGCAGATCCCTGAGCAGTCGAGTGCCAGGCCATCCACAATGTTTCATGCTAT--- 345
Qy 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102
Db 346 -----TATTCGTCATCAACAACAACAACAACAACAACATCAGCGCCAGGT 393
Qy 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
Db 394 CTCCTTCCAACAGCCTCTGCAACAATA-----TCCATTAGGCCAGGGCTCCTCCG 444
Qy 123 ThrValAsnAla 126
Db 445 GCCATCTCAGCA 456

RESULT 36
BQ244638 BQ244638 652 bp mRNA linear EST 03-MAY-2002
LOCUS TaeI5039G10R TaeI5 Triticum aestivum cDNA clone TaeI5039G10R, mRNA
DEFINITION sequence.
ACCESSION BQ244638
VERSION BQ244638.1 GI:20440514


```

|||||
Db 278 TGGAGATCACAAAGTTTGCACAAAGTACTTACAGCTGTGCAAGAATTGTGTGTCA 337
Qy 75 -----GluThrAla-----CysGlyCys 80
Db 338 GCACCTATGGCAGATCCCTGAGAGTTGCAGTGCAGCCCATCCACAAATGTGTTCATGC 397
Qy 81 ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrAlaAlaIleAsn 100
Db 398 TAT-----TATCTGCATCAACAACAACAAACAACTATCGAG 442
Qy 101 ProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMet 120
Db 443 CCAGGTCTCTTCCAAAGCTCAGCAACAATA-----TCCATTAGGCCAGGGCTC 493
Qy 121 LeuGluThrValAsnAla 126
Db 494 CTTCGGCCATCTCAGCA 511

RESULT 38
LOCUS BQ249079 657 bp mRNA linear EST 03-MAY-2002
DEFINITION TaeE25048C03R TaeE25 Triticum aestivum cDNA clone TaeE25048C03R, mRNA
sequence.
ACCESSION BQ249079
VERSION BQ249079.1 GI:20444955
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 657)
AUTHORS Cloutier,S.
TITLE Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >870 bp
Plate: 048 row: C column: 03
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
source
1..657
/cultivar="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaeE25048C03R"
/clone_lib="TaeE25"
/tissue_type="developing seeds"
/dev_stage="25 days after anthesis"
/lab_host="E. coli DH10B"
/note="vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site 1: NotI; Site 2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 25 days post-anthesis"

BASE COUNT 222 a 195 c 104 g 136 t
ORIGIN

Alignment Scores:
Pred No.: 4.51 Length: 657
Score: 87.00 Matches: 38
Percent Similarity: 37.67% Conservative: 17
Best Local Similarity: 26.03% Mismatches: 51
Query Match: 12.87% Indels: 40
DB: 14 Gaps: 7

US-10-048-196-2 (1-134) x BQ249079 (1-657)

```

```

Qy 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrPro----- 24
Db 63 GTATTTCGCAACCAACCAACCAATTTTCAGCAGCAGCAGCAGCAACAACAACA 122
Qy 25 -----GluSerAsnProLysAsnSerSerAlaAsnLeuThrThr 37
Db 123 ACACAACAACAACAACAAGAACCAANTCTTCAACAATTTTGCACAACAACAAC 182
Qy 38 SerLeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr 57
Db 183 TCCATG-----CATGGATGTTGTTATTCAGCAACAACAATAGCGCG 224
Qy 58 TrpLysIle-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer 74
Db 225 TGGAGATCACAAAGTTTTCGCAAAAGTACTTACCAGCTGTTCGAAGAATTGTGTG 284
Qy 75 -----GluThrAla-----CysGlyCys 80
Db 285 GCACCTATGGCAGATCCCTGAGAGTTGCAGTGCAGGCCATCCACAAATGTGTTCATGC 344
Qy 81 ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
Db 345 TAT-----TATCTGCATCAACAACAACAAACAACAACAACAACAACA 389
Qy 101 ProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMet 120
Db 390 CCAGGTCTCTTCCAAAGCTCAGCAACAATA-----TCCATTAGGCCAGGGCTC 440
Qy 121 LeuGluThrValAsnAla 126
Db 441 CTTCGGCCATCTCAGCA 458

RESULT 39
LOCUS BQ244737 658 bp mRNA linear EST 03-MAY-2002
DEFINITION TaeE15038C04R TaeE15 Triticum aestivum cDNA clone TaeE15038C04R, mRNA
sequence.
ACCESSION BQ244737
VERSION BQ244737.1 GI:20440613
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 658)
AUTHORS Cloutier,S.
TITLE Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 038 row: C column: 04
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
source
1..658
/cultivar="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaeE15038C04R"
/clone_lib="TaeE15"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/note="vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site 1: NotI; Site 2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 15 days post-anthesis"

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 5, 2003, 22:18:19 ; Search time 158 Seconds
(without alignments)
1909.923 Million cell updates/sec

Title: US-10-048-196-2

Perfect score: 676

Sequence: 1 MKKILYVATLMTAFTLASC.....SLKPCMLETVNAFIVPTTTR 134

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-DB=N_Geneseq_101002 -OFT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10048196 -ECCN_1_1_200 -runat_28042003_151439_5575 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6 -FGAPEXT=7
-YGAPOB=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	676	100.0	405	22	AAF59800
2	676	100.0	39003	22	AAF28534
3	118	17.5	522	21	AAZ53217
4	116	17.2	522	21	AAZ53216
5	116	17.2	102634	21	AAA81464
6	116	17.2	34980	21	AAF21609
7	116	17.2	1437668	21	AAA81490
8	110.5	16.3	519	21	AAZ53215
9	88.5	13.1	5857	19	AAV52161
10	88	13.0	6978	10	AAAN0096
11	86	12.7	3057	22	AAAD08348
12	84.5	12.5	4275	23	ABL05197
13	84.5	12.5	6959	23	ABL05196
14	84	12.4	921	23	AAAS2006
15	84	12.4	957	23	AAAS4972
16	84	12.4	1799	18	AAV74758
17	81.5	12.1	1731	21	AAZ54260
18	81	12.0	29139	23	AAAS9569
19	81	12.0	29139	23	AAAS9569
20	80	11.8	6732	23	AAK79637
21	80	11.8	6732	24	ABK84616
22	80	11.8	6798	23	AAAS68312
23	79.5	11.8	783	21	AAA82283
24	79.5	11.8	2456	24	ABK33551
25	79.5	11.8	10867	23	ABL05236
26	79.5	11.8	13329	23	AAAS71562
27	79.5	11.8	14770	24	ABL61893
28	79.5	11.8	14770	24	ABL68257
29	79.5	11.8	14770	24	ABL68538
30	79.5	11.8	14797	23	AAAS65159
31	79.5	11.8	15187	23	AAAS8131
32	79.5	11.8	69936	21	AAAF1479
33	79.5	11.8	349980	21	AAAF21607
34	79	11.7	2243	23	AAAS9666
35	79	11.7	3300	23	AAAS68171
36	79	11.7	3300	23	AAAS70249
37	79	11.7	3300	23	AAAS70795
38	79	11.7	3300	23	AAAS73830
39	79	11.7	3300	23	AAAS74991
40	79	11.7	3300	23	AAAS81666
41	78.5	11.6	600	22	AAH67310
42	78.5	11.6	34980	22	AAH68531
43	78	11.5	730	22	AAI96986
44	78	11.5	980	24	ABL34995
45	78	11.5	2617	22	AAH68570

ALIGNMENTS

RESULT 1
AAF59800
ID AAF59800 standard; DNA; 405 BP.
XX AAF59800;
XX
XX
XX 04-MAY-2001 (first entry)
XX Moraxella catarrhalis strain ATCC43617 BASB125 DNA.
XX
XX BASB125 protein; strain ATCC43617; antigen; antibody; vaccine;
KW genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory; ds.
XX
XX Moraxella catarrhalis.
XX
XX WO200109331-A2.

XX PD 08-FEB-2001.
 XX PF 27-JUL-2000; 2000WO-EP07291.
 XX PR 30-JUL-1999; 99GB-0018041.
 XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX PI Thonnard J;
 XX DR WPI; 2001-168707/17.
 XX DR P-PSDB; AAB60646.
 XX
 PT New BASB125 polypeptide isolated from Moraxella catarrhalis for
 PT treating, preventing and diagnosing diseases associated with M.
 PT catarrhalis infection in mammals, e.g. otitis media in humans
 XX
 XX Claim 13; Page 64; 73pp; English.
 XX
 CC The invention relates to the Moraxella catarrhalis strain ATCC43617
 CC BASB125 protein (AAB60646) and to DNA encoding it (AAF59800). The
 CC invention also relates to immunogenic fragments of the BASB125 protein,
 CC expression vectors and host cells comprising BASB125 nucleic acids, the
 CC recombinant production of BASB125, vaccine compositions comprising the
 CC BASB125 protein or nucleic acid, an antibody against BASB125, therapeutic
 CC compositions comprising the anti-BASB125 antibody, and a method of
 CC identifying a Moraxella catarrhalis infection via the detection of
 CC BASB125 proteins or antibodies. The vaccine compositions of the invention
 CC are useful as prophylactic or therapeutic agents against Moraxella
 CC catarrhalis infections in mammals, particularly humans. Moraxella
 CC catarrhalis is a gram negative bacterium frequently isolated from the
 CC human upper respiratory tract, which is responsible for several
 CC pathological conditions. It is responsible for about 15% of otitis media
 CC cases in children (which can lead to temporary or permanent hearing
 CC loss). It also causes pneumonia in elderly people, and sinusitis
 CC nosocomial infections and, less frequently, invasive diseases. BASB125
 CC proteins or nucleotides may additionally be used in screening for novel
 CC antibacterial compounds, and in the diagnosis and staging of infections.
 CC The present sequence represents DNA encoding the Moraxella catarrhalis
 CC strain ATCC43617 BASB125 protein.
 XX
 SQ Sequence 405 BP; 130 A; 108 C; 72 G; 95 T; 0 other;

Alignment Scores:
 Pred. No.: 5,47e-68 Length: 405
 Score: 676.00 Matches: 134
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0
 US-10-048-196-2 (1-134) x AAF59800 (1-405)
 QY 1 MetMetLysIleLeuYrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
 Db 1 ATGATGAAATTTTATATGTTACCGCACACTAATGACTGCTTTACCTTAGCAAGCTGT 60
 QY 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40
 Db 61 GCAAGTACCCCTGAAGCATTCCAAAAACAGTTCGTCTAATTTACCCACCACTTAATC 120
 QY 41 LysHISAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnYrTrpLysIle 60
 Db 121 AAACATGTCAGTCAAAACAAACCTGTCAAAACCCAGCTGCTGTCATCAGTACTGGAAAT 180
 QY 61 AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCys 80
 Db 181 GCCGCCATGAAATTTGCTTCAGAAATCCAAAGCCAAAGATTTACAGAACGGCATCGCGCTGT 240
 QY 81 ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
 Db 241 GTGGCTGATAAAGCACCTGAAGCCGCTGAGCTTGACCCAGAGTTACCAAGCTGCCATTAT 300

QY 101 ProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysHroCysMet 120
 Db 301 CCAATGCGACGACTGAAGTTGCCAAAAAATTTGGGACACTGCTTAAACCTTGATG 360
 QY 121 LeuGluThrValAsnAlaPheIleValProThrThrArg 134
 Db 361 CTAGAGACCGTCAATGCTTTTATTGTGCCAACTACCAACGC 402
 RESULT 2.
 AAF28534
 ID AAF28534 standard; DNA; 39003 BP.
 XX
 AC AAF28534;
 XX
 DT 04-APR-2001 (first entry)
 XX
 DE Genomic fragment #21.
 XX
 KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.
 XX
 OS Moraxella catarrhalis.
 XX
 PN W0200078968-A2.
 XX
 PD 28-DEC-2000.
 XX
 PF 16-JUN-2000; 2000WO-US16649.
 XX
 PR 18-JUN-1999; 99US-0140121.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PI Lagace RE, Patterson C, Berg KL;
 XX
 DR WPI; 2001-041427/05.
 XX
 PT Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids
 XX
 PS Claim 1; Page 171-180; 545pp; English.
 XX
 CC The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.
 XX
 SQ Sequence 39003 BP; 11568 A; 8751 C; 7476 G; 11208 T; 0 other;

Alignment Scores:
 Pred. No.: 2,99e-65 Length: 39003
 Score: 676.00 Matches: 134
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-10-048-196-2 (1-134) x AAF28534 (1-39003)
 QY 1 MetMetLysIleLeuYrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
 Db 6901 ATGATGAAATTTTATATGTTACCGCACACTAATGACTGCTTTACCTTAGCAAGCTGT 6960
 QY 21 AlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle 40

```

Db 6961 GCAAGTACCCCTGAAGCAATCCAAAACAGTTCCTGCTAATTTAACCCACAGCTTAATC 7020
Qy 41 LysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIle 60
Db 7021 AAACATGACGTCACAAACACCTGTCAACCCAGCTGACTGTCATCAGTACTGGAAAT 7080
Qy 61 AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCys 80
Db 7081 GCCGCGCATGAATTCCTTCAGAAATCCAAAGCCAAAGATTTTCAGAAACGCGCATGCGCTGT 7140
Qy 81 ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
Db 7141 GTGGCTGATAAAGCACCTGAAGCCGTGACGCTTGACCGAGCTTACCACAGCTGCCATTAA 7200
Qy 101 ProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMet 120
Db 7201 CCAATGACGACGACGTGAAGTGGCCCAAAAATTTGTCGACACTCGCTTAACCTTGCATG 7260
Qy 121 LeuGluThrValAsnAlaPheIleValProThrThrArg 134
Db 7261 CTAGAGACGCTCAATGCTTTTATTGTGCAACTACCACACGC 7302
RESULT 3
AAZ53217
ID AAZ53217 standard; DNA; 522 BP.
XX AC AAZ53217;
XX DT 21-MAR-2000 (first entry)
XX DE Neisseria meningitidis ORF 108 partial DNA sequence SEQ ID NO:385.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy; ds.
XX OS Neisseria meningitidis.
XX PN WO9957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX PA (CHIR ) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX PI Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
XX DR P-PSDB; AAY74455.
XX PT Novel Neisserial polypeptides predicted to be useful antigens for
XX PT vaccines and diagnostics
XX PS Claim 7; Page 320; 1453pp; English.
XX CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
XX CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ54673 represent
XX CC PCR primers used in the exemplification of the present invention. The

```

```

CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 522 BP; 134 A; 143 C; 147 G; 98 T; 0 other;

```

```

Alignment Scores:
Score: 0.000293 Length: 522
Pred. No.: 118.00 Matches: 28
Percent Similarity: 47.11% Conservative: 29
Best Local Similarity: 23.14% Mismatches: 58
Query Match: 17.46% Indels: 6
DB: 21 Gaps: 3

```

US-10-048-196-2 (1-134) x AAZ53217 (1-522)

```

Qy 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
Db 151 ATGAATAAAACCTTGTCTATT-----TTGCCGTGGCAATCTTACTCGCGGCTGC 201
Qy 21 AlaSerThrProGluSer-----AsnProLysAsnSerSerAlaAsnLeuThrThrSer 38
Db 202 GCCGCCGGGGCGGTAAACACATTCCGCAGCTTAGACGGCGGCACAGGTATGGCGGCAGC 261
Qy 39 LeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrp 58
Db 262 ATCGTCAAAATGGCGGTAGAACCAATGCCGTGGGAATTGACAAACACGACGCAATGG 321
Qy 59 LysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCys 78
Db 322 CGTTTGACCGCGCTCGCATGAGTCCGAAAACACGCGGAATGGGAAAACAGATTTCG 381
Qy 79 GlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAla 98
Db 382 GCTTGGCTCGCCCAAGACGACCAACCAACGCTGACCGGCAACGATGTG---ATGCAGATG 438
Qy 99 IleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysPro 118
Db 439 CTGGATCGCTCCACGCGCAATCAGGCACCTTGCCGCCCTGACCGCCAAAACGCTTTCCGCC 498
Qy 119 Cys 119
Db 499 TCC 501
RESULT 4
AAZ53216
ID AAZ53216 standard; DNA; 522 BP.
XX AC AAZ53216;
XX DT 21-MAR-2000 (first entry)
XX DE Neisseria meningitidis ORF 108 partial DNA sequence SEQ ID NO:383.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy; ds.
XX OS Neisseria meningitidis.
XX PN WO9957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.

```

PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 23-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR P-PSDB; AAY74454.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 PS Claim 7; Page 319; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SO Sequence 522 BP; 129 A; 142 C; 152 G; 99 T; 0 other:

Alignment Scores:
 Pred. No.: 0.000495 Length: 522
 Score: 116.00 Matches: 29
 Percent Similarity: 46.28% Conservatives: 27
 Best Local Similarity: 23.97% Mismatches: 59
 Query Match: 17.16% Indels: 6
 DB: 21 Gaps: 3

US-10-048-196-2 (1-134) x AA253216 (1-522)

QY 1 MetMetLysLeuThrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
 DB 151 ATGAATAAAACCTTGTCTATT-----TTGCCGGTGGCAATCTTACTCGGGCGTGC 201
 QY 21 AlaSerThrProGluSer-----AsnProLysAsnSerSerAlaAsnLeuThrThrSer 38
 DB 202 GCCCGCGGCGGGTAAACATTCGGCAGCTTACAGCGTGGCACAGCGTGGCGGCAGC 261
 QY 39 LeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrp 58
 DB 262 ATCGTCAAAATGGCGTGGGACCAATGCGTGGGAATTGGACAAACCCAGCAATGG 321
 QY 59 LysIleAlaMetLysLeuSerSerGluSerLysAlaLysLeSerGluThrAlaCys 78
 DB 322 CGTTTGACCGCGTGGCGATGAGTCCGAAAAACAGCGCGAGTGGGAAAAACAAGATTTCG 381
 QY 79 GlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAla 98
 DB 382 GCTTGGTCCGCCAAGAACCCGACCGGATGACCGCGCAACGATGTG---ATGCCAGATG 438
 QY 99 IleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysPro 118
 DB 439 CTGGCTCCGTCACCGCGCANTAGGCACCTTCCGCCCTGACCGCCCAACCGTTTCCGCC 498

QY 119 Cys 119
 DB 499 TGC 501
 RESULT 5
 AAA81464/c
 ID AAA81464 standard; DNA: 102634 BP.
 XX
 AC AAA81464;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_12 SEQ ID NO:12.
 XX
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW *Meningococcus B*; MenB; ds.
 XX
 OS *Neisseria meningitidis*.
 XX
 PN WO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-US233573.
 XX
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 DR WPI; 2000-318079/27.
 XX
 PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be
 PT used in the diagnosis and treatment of *N. meningitidis* infection and
 PT other *Neisseria* infections, for example, *N. gonorrhoea*.
 XX
 PS Claim 7; Page 353-383; 1760pp; English.
 XX
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA
 CC sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC *Neisseria* bacteria. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*; against all serotypes;
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 SQ Sequence 102634 BP; 23871 A; 24828 C; 27888 G; 26042 T; 5 other;

Alignment Scores:
 Pred. No.: 0.725 Length: 102634

CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

SQ Sequence 519 BP; 133 A; 142 C; 149 G; 95 T; 0 other;

Alignment Scores:

Pred. No.:	0.00208	Length:	519
Score:	110.50	Matches:	26
Percent Similarity:	45.83%	Conservative:	29
Best Local Similarity:	21.97%	Mismatches:	60
Query Match:	16.35%	Indels:	5
DB:	21	Gaps:	3

US-10-048-196-2 (1-134) x AAZ53215 (1-519)

Qy 1 MetMetLysIleLeuTyxValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
Db 151 ATGAATAAACCTTCTATT-----TTGCCGCGCGCAANTTACTCGCGGGTGC 201
Qy 21 AlaserThrProGluSer---AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
Db 202 GCCGCGCGCGCAACATTCGGCAGCTTAGACGCGCGCGGTATGGTGGCAGATC 261
Qy 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrrLys 59
Db 262 GTCAAATGAGGTGAGAAGCAATGCGGTGCGGAATTGACAGCGCGCAATGCGGT 321
Qy 60 IleAlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGly 79
Db 322 TTGACCGCGTGGCGATGAGTGCCGAAACAGCGCGGAATGGGAAACAGATTTCGCGC 381
Qy 80 CysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIle 99
Db 382 TCGGCTACCGAAGACACCTAACCGAGCTGACCGCGCAACGATGTG---ATGAGATGTG 438
Qy 100 AsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHisSerLysProCys 119
Db 439 AACCATGCCGCGCAATCAGGCACTTCCGCGCCTGACCGTCAAAACGGTTTCCGCGTGC 498

RESULT 9

AAV52161/C

ID AAV52161 standard; DNA; 5857 BP.

AC AAV52161;

DT 23-OCT-1998 (first entry)

XX Streptococcus pneumoniae genome fragment SEQ ID NO:28.

DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

XX WO9818931-A2.

PN 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19588.

XX 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX

PI

Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

Kunsch CA, Rosen CA;

XX WPI; 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae

polynucleotide sequences - useful in diagnostic kits and assays, and

pharmaceutical compositions and vaccines for Streptococcus

pneumoniae

PS Claim 1; Page 309-312; 1409pp; English.

XX The present invention describes a computer readable medium which has

the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)

recorded on it, or a representative fragment or a sequence at least 95%

identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in

SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from

Streptococcus pneumoniae. The present invention also describes an

isolated nucleic acid molecule encoding a homologue of any of the

fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the

nucleic acid molecule is produced by a process comprising: (a) screening

a genomic DNA library using as a probe a target sequence defined by any

of the sequences in SEQ ID NO:1 to 391, identifying members of the

library which contain sequences that hybridize to the target sequence and

isolating the nucleic acid molecules from the members; or (b) isolating

mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid

molecules whose nucleotide sequence is homologous to amplification

primers derived from the fragment of the S. pneumoniae genome to prime

the amplification and isolating the amplified sequences. The computer

readable medium can be used in a computer-based system for identifying

fragments of the S. pneumoniae genome of commercial importance, or

expression modulating fragments of the S. pneumoniae genome. Products

from the present invention can be used in diagnosis kits and assays, and

pharmaceutical compositions and vaccines for S. pneumoniae.

XX Sequence 5857 BP; 1800 A; 1173 C; 814 G; 2070 T; 0 other;

SQ Sequence 5857 BP; 1800 A; 1173 C; 814 G; 2070 T; 0 other;

Alignment Scores:

Pred. No.:	18.9	Length:	5857
Score:	88.50	Matches:	28
Percent Similarity:	48.54%	Conservative:	22
Best Local Similarity:	27.18%	Mismatches:	38
Query Match:	13.09%	Indels:	15
DB:	19	Gaps:	4

US-10-048-196-2 (1-134) x AAV52161 (1-5857)

Qy 30 AsnSerSerAlaAsnLeuThrThrSerLeuIleLysHis----- 42

Db 5016 AATGAAATGATACCTTACGTCACGCAAGCAAAATTAATTCATAGATGCCATGCTTACCGAG 4957

Qy 43 ---AlaValLysGlnThrCysGln-----ThrGlnLeuThrGlyHisGlnTyrr 57

Db 4956 CCACGATAGACAAAGCATGCTCAAAAGCAGGGGTATCACGCCCAACAGGTATAGATAT 4897

Qy 58 TrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAla---LysIleSerGluThr 76

Db 4896 CTAAAGTTGACGCGTTTAAAGAGCATTTGAGGATAAAGACAGATGATGATGATAA 4837

Qy 77 AlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGlu-----Leu 94

Db 4836 ACAACTCAATGCTTTTATTAGCATCTACGCTGTTTCTGTACTCAATGATATTATG 4777

Qy 95 ThrThrAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleValArgHis 114

Db 4776 ATGGACATAGGTCAACCCGTTTATAGAACTCAAGCAGCAAGGCTTACTTGAACA 4717

Qy 115 SerLeuLys 117

Db 4716 TCTTTAAG 4708

RESULT 10

AAN90096

```
ID AAN90096 standard; cDNA; 6978 BP.
XX AAN90096;
XX
XX 22-MAR-1991 (first entry)
XX
XX Sequence of plasmid pAY31 encoding wheat alpha-gliadin.
XX
XX Coeliac disease diagnosis; dough formation; food technology; ds.
XX
XX Wheat.
XX
XX US4826765-A.
XX
XX 02-MAY-1989.
XX
XX 24-OCT-1986; 86US-0922616.
XX
XX 24-OCT-1986; 86US-0922616.
XX
XX (UYHA-) UNIV OF HAWAII (USDA).
XX
XX Greene FC, Stiles JL, Nell JD, Anderson OD, Litts JC;
XX
XX WPI; 1989-150270/20.
XX
XX New wheat gluten protein gene plasmid - useful for genetic
XX transformation of yeasts to produce gliadin or glutenin
XX
XX Disclosure; Fig 4; 21pp; English.
XX
XX The preferred gene fragment in the patent of the invention encodes
XX alpha-gliadin or glutenin, and the preferred plasmid encoding alpha-
XX gliadin is pAY 31. The plasmid may be used for the prodn. of alpha-
XX gliadin, which may be used in the food technology industry to modify
XX prodn. processes of doughs and batters and to reduce unit costs.
XX Gliadin may also be used for diagnosis and treatment of illness
XX caused by wheat gluten proteins and for testing theories of dough
XX formation.
XX
XX Sequence 6978 BP; 2042 A; 1631 C; 1447 G; 1858 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 27.5 Length: 6978
XX Score: 88.00 Matches: 36
XX Percent Similarity: 37.50% Conservative: 18
XX Best Local Similarity: 25.00% Mismatches: 52
XX Query Match: 13.02% Indels: 38
XX DB: 10 Gaps: 6
XX
XX US-10-048-196-2 (1-134) x AAN90096 (1-6978)
XX
XX QY 7 ValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCysAla----- 21
XX DB 3033 GTATTGCGAACACACACACCAATTTTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCA 3092
XX
XX QY 22 -----SerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeu 39
XX DB 3093 ACAACAACAACAACAACAACAATCTTCAACAATTTTGCAACAACAATGATTCATG 3152
XX
XX QY 40 IleLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLys 59
XX DB 3153 -----CATGGATGTTGTTGTCAGCAGCAGCAGCAGCAGCAGCAGCAGTGAAG 3194
XX
XX QY 60 Ile-----AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSer----- 74
XX DB 3195 ATCACAAGTTTTCACAACAAGTACTTACCAGCTGTTGCAAGAAATTTGTTGCAACACCT 3254
XX
XX QY 75 -----GluThrAlaCysGlyCysValAla 82
XX DB 3255 ATGGCAGATCCCTGAGCAGTGCAGTGCCAGGCGATCCTCAAAGTTGTTTCATGCTAT--- 3311
XX
XX QY 83 AspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsnProAsn 102
```

```
Db 3312 -----TATTCTGCATCAACAACAACAACAACAACAACAACCATCGAGCCAGGT 3359
XX
XX QY 103 AlaArgThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGlu 122
XX DB 3360 CTCCTTCCAAACAGCCTCTGCAACAATA-----TCCATTAGGCGAGGCTCCTTCCG 3410
XX
XX QY 123 ThrValAsnAla 126
XX DB 3411 GCCATCTCAGCA 3422
XX
XX RESULT 11
XX AAD08348
XX ID AAD08348 standard; cDNA; 3057 BP.
XX
XX AC AAD08348;
XX
XX DT 09-AUG-2001 (first entry)
XX
XX DE Human secreted protein-encoding gene 4 cDNA clone HBMBD51, SEQ ID NO:14.
XX
XX KW Human; secreted protein; proliferative disorder; cancer; tumour;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX inflammation; allergy; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;
XX skin disorder; psoriasis; sepsis; diabetes; diabetes; atherosclerosis;
XX cardiovascular disorder; angiogenic disorder; kidney disorder;
XX gastrointestinal disorder; pregnancy-related disorder;
XX endocrine disorder; infection; wound healing; vulnerability;
XX cell culture; chemotaxis; food additive; gene therapy;
XX binding partner identification; ss.
XX
XX OS Homo sapiens.
XX
XX FH Key
XX CDS 413..631
XX
XX FT Location/Qualifiers
XX FT /*tag= a
XX FT /product= "Human secreted protein precursor"
XX FT /transl_except= (pos:524..526, aa:Xaa)
XX FT /note= "Xaa equals any of the naturally occurring
XX L-amino acids"
XX FT 413..496
XX FT sig_peptide
XX FT /*tag= b
XX FT 497..628
XX FT mat_peptide
XX FT /*tag= c
XX FT /product= "Mature human secreted protein"
XX
XX PN WO200077022-A1.
XX
XX XX 21-DEC-2000.
XX
XX XX 01-JUN-2000; 2000WO-US15136.
XX
XX XX 11-JUN-1999; 99US-0138629.
XX
XX XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX XX Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX WPI; 2001-367020/38.
XX P-PSDB; AAE03901.
XX
XX Nucleic acids encoding 50 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating diseases, e.g. Parkinson's
XX disease, botulism, cancers and Schmitz syndrome -
XX
XX Claim 1; Page 488-489; 614pp; English.
XX
XX AAD08345-AAD08394 represent cDNAs corresponding to 50 human secreted
XX protein genes and AAE03898-AAE03947 represent the proteins they encode.
XX AAE03948-AAE03996 represent human secreted protein fragments or variants.
XX The genes and their secreted proteins are useful for preventing,
```

CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 50 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein-encoding cDNA of the invention.

XX Sequence 3057 BP; 850 A; 909 C; 748 G; 524 T; 26 other;

Alignment Scores:
 Pred. No.: 14.9 Length: 3057
 Score: 86.00 Matches: 36
 Percent Similarity: 42.98% Conservative: 13
 Best Local Similarity: 31.58% Mismatches: 39
 Query Match: 12.72% Indels: 26
 DB: 22 Gaps: 4

US-10-048-196-2 (1-134) x AAD08348 (1-3057)

Qy 9 AlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSer---Asn 27
 Db TCTACTGTCTCCACGGCCACTAGTTTACGTTCTGTGTCAGCAGCTCCGCCAGCTCCCAAM 525
 Qy 28 ProLysAsnSerSerAlaAsn-LeuThrThrSerLeuIleLysHisAlaValLysGlnTh 47
 Db CTGCAAGCTCACCAGCAACTGTCTGTCCAGTCCCTCAAACTCCAAACCAAGCCAC 585
 Qy 47 CysGlnThrGlnLeuThrGlyHis-GlnTyrTrpLysIleAlaAlaMetLysLeuSers 67
 Db ACTCCAGGCCAGCTC-----CACCTCGGCTGGAAATGGTTATCTCTGATCCGCGCA 639
 Qy 67 erGluSerLysAlaLysIleSerGluThrAlaCysGlyCys----- 80
 Db GCAGTGACAGTGGCGGGTTCAGCGTCAGGGCTGTGGCTGTGCCAGCTCTGACATGTCT 699
 Qy 81 -----ValAlaAspLys-----Alap 86
 Db CCAGCAGRACAGNTCAACAGATGGCTGCACAGCAGCAACAAAGGGCCAAACTCATGCG 759
 Qy 86 roGluAlaValSerLeuThrGluLeuThrThrAlaAla 98
 Db CAGAGCAGCAGCACACACAGCAGCAGCAGCAGCAGCA 797

RESULT 12

ABL05197
 ID ABL05197 standard; cDNA: 4275 BP.

XX ABL05197;

AC ABL05197;
 XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 10073.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW

pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

XX P-PSDB; ABB61094.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Claim 1; SEQ ID NO 10073; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
 CC sequences (ABLI01840-ABLI6175) and the encoded proteins
 CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4275 BP; 1189 A; 1310 C; 1063 G; 713 T; 0 other;

Alignment Scores:

Pred. No.: 35 Length: 4275
 Score: 84.50 Matches: 31
 Percent Similarity: 39.06% Conservative: 19
 Best Local Similarity: 24.22% Mismatches: 32
 Query Match: 12.50% Indels: 46
 DB: 23 Gaps: 3

US-10-048-196-2 (1-134) x ABL05197 (1-4275)

Qy 13 ThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSer 32
 Db ACCACAACAAGTGTGAGCTCATCAGCAACACAGCAGCTGCACCAACCACTCCAGT 1403
 Qy 33 AlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeu 52
 Db GCAACACCCACACACAGCTGCCCTCCGCCAGCGCTTTTGGTTCAGCTCTCGATTCGCCAG 1463
 Qy 53 ThrGlyHisGln-TyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAla 72
 Db TCCGCCACAGCTAC----- 1479
 Qy 72 sileSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuTh 92
 Db -----GC 1481
 Qy 92 rGluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGlu----- 106
 Db CGAATCCACCACTTCCATCAGAGCTTGGCCCGAGTGGAGAGATGAAGTCTCTGA 1541
 Qy 107 -ValAlaGlnLysIleValArgHisSerLeu-----LysProCysMetLe 121
 Db CTTAAGCAGCATCTCAAGCCGAGGAGACTTGGCCGTTTCGGTTCCTAAACACCATCTGAT 1601

QY 121 uGluThrValAsnAlaPheille 128
 ID ABL05196 standard; cDNA; 6959 BP.
 DB 1602 TGAGAGGCTGAACCGTATCTG 1623

RESULT 13
 ABL05196/c
 AC ABL05196;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10070.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE-CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR P-PSDB; ABB61093.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Claim 1; SEQ ID NO 10070; 21pp + Sequence Listing; English.
 PS
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6959 BP; 1517 A; 1529 C; 1832 G; 2081 T; 0 other;

Alignment Scores:
 Pred. No.: Length: 68.6 6959
 Score: 84.50 Matches: 31
 Percent Similarity: 39.06% Conservative: 19
 Best Local Similarity: 24.22% Mismatches: 32
 Query Match: 12.50% Indels: 46
 DB: 23 Gaps: 3

US-10-048-196-2 (1-134) x ABL05196 (1-6959)

QY 13 ThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSer 32
 DB 3932 ACCAACACACACCTTCAGCTGCATCAGCAACACACAGCAGCTGCACCACTCCAGT 3873

QY 33 AlaAsnLeuThrThrSerLeuLeuLeuHisAlaValLysGlnThrCysGlnThrGlnLeu 52
 DB 3872 GCACACCCACCACTCCCTCCGCGAGGCTTTCGGTCAGCTCTCGATTCGCCCCAG 3813

QY 53 ThrGlyHisGln-TyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLys 72
 DB 3812 TCCGGCCACCAAGCTAC----- 3797

QY 72 sIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuTh 92
 DB 3796 -----GC 3795

QY 92 rGluLeuThrThrAlaAlaIleAsnProAsnAlaAArgThrGlu----- 106
 DB 3794 CGAATCCACCACTCCATCAGACTTGGCCCGACTGGAGAAGATGAAGTCTCTGA 3735

QY 107 -ValAlaGlnLysIleValArgHisSerLeu-----LysProCysMetLe 121
 DB 3734 CCTAAGCAGCATCTCAAGCGCAGGAACTTCCCGTTCCGGTCTTAACCACTCATGAT 3675

QY 121 uGluThrValAsnAlaPheille 128
 DB 3674 TGAGAGGCTGAACCGTATCTG 3653

RESULT 14
 AAS52006/c
 ID AAS52006 standard; DNA; 921 BP.
 XX
 AC AAS52006;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Staphylococcus aureus DNA for cellular proliferation protein #423.
 XX
 KW Antisense; ds; prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Staphylococcus aureus.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-OCT-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR P-PSDB; AAU34147.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Claim 27; Seq ID No 4588; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 921 BP; 322 A; 135 C; 189 G; 275 T; 0 other;

Alignment Scores:

Pred. No.: 4.79 Length: 921
Score: 84.00 Matches: 34
Percent Similarity: 45.30% Conservative: 19
Best Local Similarity: 29.06% Mismatches: 44
Query Match: 12.43% Indels: 20
DB: 23 Gaps: 6

US-10-048-196-2 (1-134) x AAS52006 (1-921)

Qy 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAla-----PheThr 16
Db 539 ATGCCACGAGTACAGTACGTAAACC---ACTTTTATCTGCAATAAGCGTTTATCTCT 483
Qy 17 LeuAlaSerCysAlaSerThr-----ProGluSer 26
Db 482 TTAATCCAGTCAGGCAAAATCTCTAAAGCGCGTGATATTGGACGAATTCACCTCGGAAA 423
Qy 27 AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGln 46
Db 422 TGACCTAAATCAAAATTCATAATCATTAATCATTAATCATTAATCATTAATCATTAAT 363
Qy 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMet-----Lys 64
Db 362 AGAGCTTTCTCTAAATCTACAGGTGATAAATATTGGCCAGTGTGTCTTGGATCGAGC 303
Qy 65 LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLys 84
Db 302 TCTTTCTTCTAAATCTAAAGCAACAATCTCTTTTCGAGGACGTACATGATCTTTCTTAAA 243
Qy 85 Ala---ProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
Db 242 GCATGTCCTTCAGCTTCATCAATT-----TTAAACACCATATCTTTGAAT 198

RESULT 15

AAS54972/C
ID AAS54972 standard; DNA; 957 BP.

XX
XX
AC AAS54972;

XX
DT -13-FEB-2002 (first entry)

XX Staphylococcus aureus DNA for cellular proliferation protein #1284.

XX Antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.

XX Staphylococcus aureus.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR P-PSDB; AAU37113.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 27; Seq ID No 8609; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 957 BP; 335 A; 141 C; 194 G; 287 T; 0 other;

Alignment Scores:

Pred. No.: 5.05 Length: 957
Score: 84.00 Matches: 34
Percent Similarity: 45.30% Conservative: 19
Best Local Similarity: 29.06% Mismatches: 44
Query Match: 12.43% Indels: 20
DB: 23 Gaps: 6

US-10-048-196-2 (1-134) x AAS54972 (1-957)

Qy 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAla-----PheThr 16
Db 542 ATGCCACGAGTACAGTACGTAAACC---ACTTTTATCTGCAATAAGCGTTTATCTCT 486

Qy 17 LeuAlaSerCysAlaSerThr-----ProGluSer 26

Db 485 TTAATCCAGTCAGGCAAAATCTCTAAAGCGCGTGATATTGGACGAATTCACCTCGGAAA 426

Qy 27 AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGln 46

Db 425 TGACCTAAATCAAAATTCATAATCATTAATCATTAATCATTAATCATTAATCATTA 366

Qy 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMet-----Lys 64

Db 365 AGAGCTTTCTCTAAATCTACAGGTGATAAATATTGGCCAGTGTGTCTTGGATCGAGC 306

Qy 65 LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLys 84

Db 305 TCTTCTTCTAAATCTAAAGCAACAATCTCTTTTCGAGGACGTACATGATCTTTCTTAAA 246

Qy 85 Ala---ProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100

Db 245 GCATGTCCTTCAGCTTCATCAATT-----TTAAACACCATATCTTTGAAT 201

RESULT 16

AAV74758/c

AAV74758 standard; DNA; 1799 BP.
 AAV74758;
 16-MAR-1999 (first entry)
 Staphylococcus aureus contig SEQ ID #447.
 DE
 XX
 XX
 KW Computable readable medium; vaccine; S. aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 XX
 XX Staphylococcus aureus.
 XX
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1141..1200
 FT /tag= a
 FT /note= "these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"
 XX
 XX
 PN EP786519-A2.
 XX
 XX 30-JUL-1997.
 XX
 XX 07-JAN-1997; 97EP-0100117.
 XX
 XX 05-JAN-1996; 96US-0009861.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 PI Rosen CA;
 XX
 XX WPI: 1997-374922/35.
 XX
 XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
 stored on computer readable medium and used in the production of
 anti-S. aureus vaccines
 XX
 XX Claim 1: Page 1360-1361; 3271pp; English.
 XX
 XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 of the invention. The DNA sequences are recorded on a computer readable
 medium, preferably selected from a floppy or hard disk, random access
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 the S. aureus DNA sequences allows putative functions to be assigned so
 that protein-encoding or regulatory regions of commercial, therapeutic or
 industrial importance can be obtained. Specifically, sequences which are
 likely to encode antigens have been identified and these polypeptides can
 be used in a vaccine composition against S. aureus infection. The
 polypeptides can also be used in a kit for the immunodetection of
 S. aureus in a sample. S. aureus is implicated in numerous human diseases,
 including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 skin and surgical wound infections, scalded skin syndrome, toxic shock
 syndrome, etc. Organisms transformed with the DNA sequences can be used
 for recombinant production of the polypeptides. The new DNA sequences
 (and their fragments) are useful as primers or probes for isolating
 homologues of any of the S. aureus DNA sequences contained on the
 computer readable medium.
 XX
 XX Sequence 1799 BP; 601 A; 280 C; 320 G; 533 T; 65 other;

Alignment Scores:

Pred. No.:	12.1	Length:	1799
Score:	84.00	Matches:	34
Percent Similarity:	45.30%	Conservative:	19
Best Local Similarity:	29.06%	Mismatches:	44
Query Match:	12.43%	Indels:	20
DB:	18	Gaps:	6

US-10-048-196-2 (1-134) x AAV74758 (1-1799)
 QY 1 MetMetLysIleLeuTyrValThrAlaThrLeuMetThrAla-----PheThr 16
 DB 663 ATGCCACCAGTACAGTACGTAACC---ACTTTTATCTGCAAAATACGCTTTATTCTCT 607
 QY 17 LeuAlaSerCysAlaSerThr-----ProGluSer 26
 DB 606 TTAATCCAGTCAGCCAAATCTCTAAACGTTGTGATATTTGGACGAATTCACCTCCGAAA 547
 QY 27 AsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGln 46
 DB 546 TGACCTAAATCAAAATTCATAATCATTACGTGCATCAATAATGACTGTGTCATCATCTTCA 487
 QY 47 ThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMet-----Lys 64
 DB 486 AGAGCTTTTCTAAATCTCAGGTGATAAATATTTGCCAGTGTGTCTTGGATCGACG 427
 QY 65 LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLys 84
 DB 426 TCATCTTCTAAATCTAAAGCAACGATTCTTTTCGAGGAGGTACATGCATTTTCTTAAAA 367
 QY 85 Ala---ProGluAlaValSerLeuThrGluLeuThrThrAlaAlaIleAsn 100
 DB 366 GCATGCTCTTCAGCTTCATCAATT-----TTAAACACCATATCTTTGAAT 322
 RESULT 17
 AAZ54260
 ID AAZ54260 standard; DNA; 1731 BP.
 XX
 AC AAZ54260;
 XX
 DT 21-MAR-2000 (first entry)
 DE
 DE Neisseria gonorrhoeae ORF 730 partial DNA sequence SEQ ID NO:2469.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 OS
 OS Neisseria gonorrhoeae.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999;
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masiagnani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR P-PSDB; AAY75498.
 XX
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 PS Claim 7: Page 1181; 1453pp; English.
 XX
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941

CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present inventions. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC *Neisseria* bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX Sequence 1731 BP; 507 A; 521 C; 418 G; 285 T; 0 other;

Alignment Scores:
Pred. No.: 22.1 Length: 1731
Score: 81.50 Matches: 30
Percent Similarity: 36.43% Conservative: 17
Best Local Similarity: 23.26% Mismatches: 45
Query Match: 12.06% Indels: 37
DB: 21 Gaps: 3

US-10-048-196-2 (1-134) x AA254260 (1-1731)

QY 17 LeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThr 36
DQ 275 TTGTTTACACCGTCGGCTTTCCGGACACGGACACGAAGAACACGCCCTTCGACAACC 334
QY 37 ThrSerLeuLeuLysHisAlaValLysGlnThrCysGlnThrGlnLeuThrGlyHisGln 56
DQ 335 ACGCGCGGACAGCGCAAGCAAGAAAGCAACGTTGACG----- 376
QY 57 TyrTrpLysLeuAlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThr 76
DQ 377 -----ACG 379
QY 77 AlacysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThr 96
DQ 380 GCTTTACGGTGATACCGGCTCAACTGGGAAGGACACGAACATCATCCCGCGCATCCCTACG 439
QY 97 -----AlaAlaIleAsnProAsnAlaArg----- 104
DQ 440 ACGCGCGGACAGCGCGCAATTTACCCCAACCTACGGCGGACGACGACGATACCTATC 499
QY 105 ThrGluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThrVal 124
DQ 500 ACGTCAACGGCACACCGCGCAGTATCAAACTCAATCCGACCGACACCGCGCATCCGGC 559
QY 125 AsnAlaPheIleValProThrThrThr 133
DQ 560 AACGCATATTGACAACTACACAACCC 586

RESULT 18

AA559569
ID AA559569 standard; DNA; 29139 BP.

XX AA559569;

DT 13-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein encoding DNA #64.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant; ds.

OS Propionibacterium acnes.

XX WO200181581-A2.

PN 01-NOV-2001.

PD

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris

XX Claim 1; SEQ ID No 64; 1069pp; English.

XX Sequences AA559506-AA559804 represent DNA molecules encoding
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
CC associated DNA sequences are used in the treatment, prevention and
CC diagnosis of medical conditions caused by *P. acnes*. The disorders include
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC osteomyelitis), uveitis and endophthalmitis. *P. acnes* is also involved
CC in infections of bone, joints and the central nervous system, however it
CC is particularly involved in the inflammatory lesions associated with acne
CC vulgaris. A method for detecting the presence or absence of *P. acnes* in a
CC patient comprises contacting a sample with a binding agent that binds to
CC the proteins of the invention and determining the amount of bound protein
CC in the sample. The polypeptides may be used as antigens in the production
CC of antibodies specific for *P. acnes* proteins. These antibodies can be
CC used to downregulate expression and activity of *P. acnes* polypeptides and
CC therefore treat *P. acnes* infections. The antibodies may also be used as
CC diagnostic agents for determining *P. acnes* presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC polypeptides shown in AAU5305-AAU53603 and AAU67567-AAU67570.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 29139 BP; 6272 A; 9455 C; 8116 G; 5292 T; 4 other;

Alignment Scores:

Pred. No.: 1.24e+03 Length: 29139
Score: 81.00 Matches: 41
Percent Similarity: 44.44% Conservative: 23
Best Local Similarity: 28.47% Mismatches: 58
Query Match: 11.98% Indels: 22
DB: 23 Gaps: 6

US-10-048-196-2 (1-134) x AA559569 (1-29139)

QY 3 LysIleLeuTyrValThr-----AlaThrLeuMetThrAlaPhe 15

DQ 1471 AAATTGGTGTACCAGACGTTGAGTACCCATGAACAACGACATTCGCTCGCGCTCG 1530

QY 16 ThrLeuAlaSerCysAlaSerThrProGluSerAsnProLysAsnSerSerAlaAsnLeu 35

DQ 1531 GCAGTGGCGAAGTCTCTCTGACG---CGGTGAATCCGGCAGGAGTCTCTCGAACTC 1587

QY 36 ThrThrSerLeuIleLysHisAlaValLysGlnThrCysGln----- 49

DQ 1588 TTCGGCCGATCGCGACGGCAAGCGACGTACCGAGCGGAATCCACCGAGTATCGGCCA 1647

QY 50 ---ThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGlu 68

DQ 1648 CCCACCCAGCTCCAGAACCCAAAGCGGTTCTACTGGGTGCGATCGCGAATCTCTGACCTTG 1707

QY 69 SerLysAlaLysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAla 88

DQ 1708 TCCAAAGGCC-----GTGAGACGCGCGCAAAAGTCTTTAGGACGCGGCTCAGCTCGAGCC 1761

QY 89 ValSerLeuThrGluLeuThrThrAlaAlaIleAsnPro-----AsnAlaArgThr 105
 DB 1762 TGCTCGTCTCGCAATCGCACCGGAGCAGCAACAGCCGCTCAAGCAGCAGCACGACGACG 1821
 QY 106 GluValAlaGlnLysIleValArgHisSerLeuLysProCysMetLeuGluThrValAsn 125
 DB 1822 CACCGGGCA-----TTGTCAGAGTCTCCAAAGTGGTGAAGGTCTTAGAGGCAATGATG 1875
 QY 126 AlaPheIleVal 129
 DB 1876 ACCAGGTCGTC 1887
 RESULT 19
 ID AAS59569/c
 XX AAS59569 standard; DNA; 29139 BP.
 AC AAS59569;
 XX
 DT 13-FEB-2002 (first entry)
 DE Propionibacterium acnes immunogenic protein encoding DNA #64.
 XX
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant; ds.
 XX
 OS Propionibacterium acnes.
 XX
 PN WO200181581-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001WO-US12865.
 XX
 PR 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX
 DR WPI; 2001-616774/71.
 XX
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX
 PS Claim 1; SEQ ID No 64; 1069pp; English.
 XX
 CC Sequences AAS59506-AAS59804 represent DNA molecules encoding
 CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
 CC associated DNA sequences are used in the treatment, prevention and
 CC diagnosis of medical conditions caused by P. acnes. The disorders include
 CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
 CC in infections of bone, joints and the central nervous system, however it
 CC is particularly involved in the inflammatory lesions associated with acne
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a
 CC patient comprises contacting a sample with a binding agent that binds to
 CC the proteins of the invention and determining the amount of bound protein
 CC in the sample. The polypeptides may be used as antigens in the production
 CC of antibodies specific for P. acnes proteins. These antibodies can be
 CC used to downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
 CC polypeptides shown in AAU55305-AAU55603 and AAU67567-AAU67570.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 29139 BP; 6272 A; 9455 C; 8116 G; 5292 T; 4 other;
 Alignment Scores:
 Pred. No.: 1.24e+03 Length: 29139
 Score: 81.00 Matches: 31
 Percent Similarity: 47.15% Conservative: 27
 Best Local Similarity: 25.20% Mismatches: 41
 Query Match: 11.98% Indels: 24
 DB: 23 Gaps: 6
 US-10-048-196-2 (1-134) x AAS59569 (1-29139)
 QY 12 MetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnProLys----- 29
 DB 19713 CTCACGATTATAGATTAGGCTCTCTCGAATTCTCAGATCACTGGCGATATTGGGGCACACTCT 19594
 QY 30 -----AsnSerSerAlaAsnLeuThrThrSerLeuIleLysHisAla 43
 DB 19653 CCGATATTCTGGGAGAGTTCTTCGAATTCTCAGATCACTGGCGATATTGGGGCACACTCT 19594
 QY 44 ValLysGlnThrCysGln-----ThrGlnLeuThrGlyHisGlnTrpTrpLysIle 60
 DB 19593 GTCACCTGGTTCGTCGCCAGTTGTGGTCATCATGTCAGGTTGGCTATGAGGAAAAGATGAGA 19534
 QY 61 AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCys 80
 DB 19533 GCACCAATGAGGATT-----GCTGAGACGACACGACGACGTTGCCACGGGAGC 19483
 QY 81 -----ValAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThr 95
 DB 19482 AACCGTTTAAGGAACCCAGCACCAAGGCCCGCCCTCCATAGCCATCAGC-----TTG 19429
 QY 96 ThrAlaAlaIleAsnProAsnAlaArgThr-----GluValAlaGlnLysIleVal 112
 DB 19428 ACGTCAACGATCGTCCGATCGCCAGAGGCGAGCCAGTGGCGGTGGGAGAACTCGTC 19369
 QY 113 ArgHisSer 115
 DB 19368 AGAGAAACA 19360
 RESULT 20
 ID AAS79637/c
 XX AAS79637 standard; cDNA; 6732 BP.
 AC AAS79637;
 XX
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #15441.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG15450.
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 15441; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6732 BP; 1512 A; 2090 C; 1685 G; 1445 T; 0 other;

Alignment Scores:
Pred. No.: 213 Length: 6732
Score: 80.00 Matches: 24
Percent Similarity: 46.32% Conservative: 20
Best Local Similarity: 25.26% Mismatches: 36
Query Match: 11.83% Indels: 15
DB: 23 Gaps: 3

US-10-048-196-2 (1-134) x AAS79637 (1-6732)

Qy 15 PheThrLeuAlaSerCysAlaSerThrProGlu----- 25
Db 2128 TTTAGGATGGCGGTGTGGACCTCCATCCACGTCGACGAGCAATCAGGAGTTC 2069
Qy 26 ---SerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuLeuLysHisAlaVal 44
Db 2068 ATCACTTCACCCCGCTCCAGGAAGCGTCTCTCTGGTGAATTG---CGCACTGCTGTG 2012
Qy 45 LysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLys 64
Db 2011 AGTGTGCTCTGCACAATACCCATGACAGCGCGATTGCTCTGGGGGTGCACATCATGCGA 1952
Qy 65 -----LeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGlyCysVa 81
Db 1951 GGAACATGGCCAGCTCTGATCTCTCTCGTCTCCAGAGATGTGGCAGGTGCGAAG 1892
Qy 81 LalaAspLysAlaProGluAlaValSerLeuThrGluLeuThr 95
Db 1891 TTCACTCATCCCCGTCAAAGTCTGCATTGTACGGAGTTGTCA 1849
RESULT 21
ABK84616/c
ID ABK84616 standard; cDNA; 6732 BP.
XX
AC ABK84616;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1187.
XX
XX Human; ss: granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;

KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
PA (GENE-) GENE LOGIC INC.
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity
PS Claim 1; SEQ ID No 1187; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from GS, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from GS in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6732 BP; 1512 A; 2090 C; 1685 G; 1445 T; 0 other;

Alignment Scores:
Pred. No.: 213 Length: 6732
Score: 80.00 Matches: 24
Percent Similarity: 46.32% Conservative: 20

QY 1 MetMetLysIleLeuYrValThrAlaThrLeuMetThrAlaPheThrLeuAlaSerCys 20
Db 1756 TTGAGGAGTCTTCTTCTCAGAGCTACGTTAAATATGTATACAAAGCTGTCAAGCTGT 1815
QY 21 AlaSerThrProGluSer- 26
Db 1816 GGTTCATGCCAGAGCGGGTTTCAGCAAGCTGCAGATCTGATTGATGCTGAGCAACGAA 1875
QY 27 -----AsnProLysAsnSerAlaAsnLeuThr-----Ser-LeuIlely 41
Db 1876 TGAAGAAGTCCATGTTGGGTCAAGTCTGCTGCACAGAGGTTCTTCAATACTTAT 1935
QY 41 sHisAlaValLys-----GlnThrCysGlnThrGlnLeuThrGlyHisGlnTyrTrpLy 59
Db 1936 GCATGACATCCCAAGTTAAAGGGTTGTGCAACTAGCTGAGAGGAAATCAAGAATGGAA 1995
QY 59 stle-AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysG 79
Db 1996 AATGTGTTGAATGCTGTCAGCTCTACAGGAGAGCTAGA----- 2036
QY 79 LysValAlaAlaAspLysAlaProGluAlaValSerLeuThrGluLeuThrThrAlaAlaI 99
Db 2037 -----ACATTAGAACCTTTGGAAGAGCGGGG 2064
QY 99 leAsnProhsnAlaArgThrGluValAlaGlnLysIleValArg-HisSerLeuLysPro 118
Db 2065 GAGAAATTAATGTTTGTTCACCTGCGCAAGGTTGTTGCAGTCACTCATTTGAAAAAC 2124
QY 119 Cys---MetLeuGluThrValAsnAlaPheIleVal 129
Db 2125 ATTTCTCTCCAGACAGAAAAAACTTTATAGTT 2160
RESULT 25
ID ABL05236
AC ABL05236;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10190.
XX
KW Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
DR P-PSDB; ABB61133.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 10190; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72022).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 10867 BP; 3008 A; 2450 C; 2315 G; 3094 T; 0 other;
Alignment Scores: 471 Length: 10867
Pred. No.: 79.50 Matches: 19
Score: 45.45% Conservative: 11
Percent Similarity: 28.79% Mismatches: 29
Best Local Similarity: 11.76% Indels: 7
Query Match: 23 Gaps: 2
DB: 2
US-10-048-196-2 (1-134) x ABL05236 (1-10867)
QY 31 SerSerAlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGlnThrCysGlnThr 50
Db 4234 ACAACAGCGAATTACACACGGGAAACACACACACACAACTGAGGAAACATGAAACA 4293
QY 51 GlnLeuThrGlyHisGlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLys 70
Db 4294 CAA-----TGGAAACAGCAGCAGCAGCGGGGAAAAATCCGGAACAAA 4335
QY 71 ---AlaLysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaVal 89
Db 4336 GGGCGGAGGATCGAGGAGTCTCTGTGGGTAAATTACGACAAAGTGCACACATTAGCT 4395
QY 90 SerLeuThrGluLeuThr 95
Db 4396 TCCGCAACAGCTTTGTCC 4413
RESULT 26
ID AAS71562 standard; cDNA; 13329 BP.
XX
XX AAS71562;
AC AAS71562;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #7366.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG07375.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID No 7366; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 13329 BP; 4337 A; 3011 C; 2943 G; 3038 T; 0 other;

Alignment Scores:

Pred. No.:	624	Length:	13329
Percent Similarity:	79.50	Matches:	28
Best Local Similarity:	43.81%	Conservative:	18
Query Match:	26.67%	Mismatches:	52
	11.76%	Indels:	7
		Gaps:	4

US-10-048-196-2 (1-134) x AAS71562 (1-13329)

Qy	14	AlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro---LysAsnSerSer	32
Db	11896	GCCTTACTACTCTTCTGATGTAGATGTAAGTCCAGATTCCAGTGAACACACACACCC	11955
Qy	33	AlaAsnLeuThrThrSerLeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeu	52
Db	11956	AGGGAT-----AACATAATT-----GCAGTAGAAGAGCATGTGCCACACAAAG	12000
Qy	53	ThrGlyHisGlnThrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLys	72
Db	12001	CAAGGG---CAGCCAGAGAAAGGCAAGCCAAACAGCTTCATCCAAAGTCCCAAGTAAG	12057
Qy	73	IleSerGluThrAlaCysGlyCysValAlaAlaAspLysAlaProGluAlaValSerLeuThr	92
Db	12058	GTAAGATCCACCTGTGTCTACTACACCCACCACCTGCCACCCACCCACCTACACCC	12117
Qy	93	GluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleVal	112
Db	12118	ACTACCCACCCAGCTGCACAGTAAAGTTAGGAAAGTCAGCTCAAGGAAGTATGT	12177
Qy	113	ArgHisSerLeuLys 117	
Db	12178	AAACATTCATTGAA 12192	

RESULT 27

ABL61893	
ID	ABL61893 standard; DNA; 14770 BP.
XX	
AC	ABL61893;
XX	
DT	15-MAY-2002 (first entry)
XX	
DE	Colon adenocarcinoma related gene sequence SEQ ID NO:230.
XX	
KW	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW	cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KW	gene; ds.

XX OS Homo sapiens.
XX PN WO200194629-A2.
XX PD 13-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US10838.
XX OS 05-JUN-2000; 2000US-209473P.
XX PR 05-JUN-2000; 2000US-209531P.
XX PR 18-SEP-2000; 2000US-233133P.
XX PR 18-SEP-2000; 2000US-233617P.
XX PR 20-SEP-2000; 2000US-234009P.
XX PR 20-SEP-2000; 2000US-234034P.
XX PR 20-SEP-2000; 2000US-234052P.
XX PR 22-SEP-2000; 2000US-234509P.
XX PR 22-SEP-2000; 2000US-234567P.
XX PR 23-SEP-2000; 2000US-234923P.
XX PR 23-SEP-2000; 2000US-234924P.
XX PR 25-SEP-2000; 2000US-235077P.
XX PR 25-SEP-2000; 2000US-235082P.
XX PR 25-SEP-2000; 2000US-235134P.
XX PR 25-SEP-2000; 2000US-235280P.
XX PR 26-SEP-2000; 2000US-235637P.
XX PR 26-SEP-2000; 2000US-235638P.
XX PR 27-SEP-2000; 2000US-235711P.
XX PR 27-SEP-2000; 2000US-235720P.
XX PR 27-SEP-2000; 2000US-235840P.
XX PR 27-SEP-2000; 2000US-235863P.
XX PR 28-SEP-2000; 2000US-236028P.
XX PR 28-SEP-2000; 2000US-236032P.
XX PR 28-SEP-2000; 2000US-236033P.
XX PR 28-SEP-2000; 2000US-236034P.
XX PR 28-SEP-2000; 2000US-236109P.
XX PR 28-SEP-2000; 2000US-236111P.
XX PR 29-SEP-2000; 2000US-236842P.
XX PR 29-SEP-2000; 2000US-236891P.
XX PR 02-OCT-2000; 2000US-237172P.
XX PR 02-OCT-2000; 2000US-237173P.
XX PR 02-OCT-2000; 2000US-237278P.
XX PR 02-OCT-2000; 2000US-237294P.
XX PR 02-OCT-2000; 2000US-237295P.
XX PR 02-OCT-2000; 2000US-237316P.
XX PR 03-OCT-2000; 2000US-237425P.
XX PR 03-OCT-2000; 2000US-237598P.
XX PR 03-OCT-2000; 2000US-237604P.
XX PR 03-OCT-2000; 2000US-237606P.
XX PR 03-OCT-2000; 2000US-237608P.
XX PR 01-NOV-2000; 2000US-244867P.
XX PR 01-NOV-2000; 2000US-245084P.

(AVAL-) AVALON PHARM.

XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;

XX WPI; 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set

XX Claim 1; SEQ ID 230; 44pp; English.

XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening

CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.
 XX
 SQ Sequence 14770 BP; 4829 A; 3240 C; 3211 G; 3490 T; 0 other;

Alignment Scores:

Pred. No.: 719 Length: 14770
 Score: 79.50 Matches: 28
 Percent Similarity: 43.81% Conservative: 18
 Best Local Similarity: 26.67% Mismatches: 52
 Query Match: 11.76% Indels: 7
 DB: 24 Gaps: 4

US-10-048-196-2 (1-134) x ABL61893 (1-14770)

QY 14 AlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro---LysAsnSerSer 32
 DB 11890 GCCCTTACTACTCTCTCATGTGTAGATGTAAGTCCAGATTCCAGTGAACACACACACC 11949
 QY 33 AlaAsnLeuThrSerLeuIleLysHisAlaValLysGlnThrCysGlnThrGlnLeu 52
 DB 11950 AGGGAT-----AACATAATT-----GCAGTATGAAAAGCATGTGCCACACAAAG 11994
 QY 53 ThrGlyHisGlnThrTrpLysIleAlaMetLysLysSerSerGluSerLysAlaLys 72
 DB 11995 CAAGGG--CAGCCAGAGAAGGCAAGGCAACAGCTTCCATCCAAAGTGCAGTAAAG 12051
 QY 73 IleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThr 92
 DB 12052 GTAAGATCACTGTGCTACTACCACCACCACCTGCTCCACACACACCTACCTACCACC 12111
 QY 93 GluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleVal 112
 DB 12112 ACTACCACCACCACCTGTCAGTAAAGTTAGAAAAGTCAGCTCAAGGAAGTATGT 12171
 QY 113 ArgHisSerLeuLys 117
 DB 12172 AAACATTCATTGAA 12186

RESULT 28

ID ABL68257
 XX ABL68257 standard; DNA; 14770 BP.
 AC ABL68257;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Kidney cancer related gene sequence SEQ ID NO:6594.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-US10838.
 XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-233133P.
 PR 18-SEP-2000; 2000US-233617P.

PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235637P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 XX
 XX (AVAL-) AVALON PHARM.

Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 Soppet DR, Weaver Z;

WPI: 2002-188264/24.

Screening for anti-neoplastic agent involves exposing cells to a
 chemical agent to be tested for anti-neoplastic activity, and
 determining a change in expression of a gene of a signature gene set -

Claim 1; SEQ ID 6594; 44pp; English.

The present invention describes a method (M1) for screening for an
 anti-neoplastic agent. The method involves exposing cells to a chemical
 agent to be tested for anti-neoplastic activity, determining a change in
 expression of at least one gene (I) of a signature gene set, where (I)
 comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 to ABL70110), or is at least 95% identical to (S), where a change in
 expression is indicative of anti-neoplastic activity. (I) has cytostatic
 activity and can be used in gene therapy. M1 can be used for screening
 an anti-neoplastic agent, and can be used for producing a product which
 is the data collected with respect to the anti-neoplastic agent as a
 result of M1, and the data is sufficient to convey the chemical
 structure and/or properties of the agent. M1 can be used in the
 treatment of cancer such as colon, breast, stomach, lung, thyroid,
 oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.

Sequence 14770 BP; 4829 A; 3240 C; 3211 G; 3490 T; 0 other;

Alignment Scores:


```
Pred. No.: 719 Length: 14770
Score: 79.50 Matches: 28
Percent Similarity: 43.81% Conservative: 18
Best Local Similarity: 26.67% Mismatches: 52
Query Match: 11.76% Indels: 7
DB: 24 Gaps: 4

US-10-048-196-2 (1-134) x ABL68538 (1-14770)

QY 14 AlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro---LysAsnSerSer 32
DB 11890 GCCCTTACTTCTTCTTCATGCTAGATGTAAGTCCAGATTCAGTGAACACACACCC 11949

QY 33 AlaAsnLeuThrThrSerLeuLeuLysHisAlaValLysGlnThrCysGlnThrGlnLeu 52
DB 11950 AGGGAT-----AACATAATT-----GCAGTTAGAAAAGCATGTGCCACACAAAAG 11994

QY 53 ThrGlyHisGlnTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLys 72
DB 11995 CRAAGG---CAGCAGAGAAAGCAAGGCCAACACGCTCCATCCCAAGTTGCCAGTAAAG 12051

QY 73 IleSerGluThrAlaCysGlyCysValAlaAlaAspLysAlaProGluAlaValSerLeuThr 92
DB 12052 GTAAGATCCACCTGTGTCACCTACCACCACCACCTGCCACACCACCACTACCACCC 12111

QY 93 GluLeuThrThrAlaAlaIleAsnProAsnAlaAlaArgThrGluValAlaGlnLysIleVal 112
DB 12112 ACTACACCACCACCACCTGCACAGTAAAGTTAGGAAAAGTCAGCTCAAGGAAGTATGT 12171

QY 113 ArgHisSerLeuLys 117
DB 12172 AACATTCATTGAA 12186

RESULT 29
ABL68538
ID ABL68538 standard; DNA; 14770 BP.
XX
AC ABL68538;
XX
DT 15-MAY-2002 (first entry)
XX
DE Kidney cancer related gene sequence SEQ ID NO:6875.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.

27-SEP-2000; 2000US-235711P.
27-SEP-2000; 2000US-235720P.
27-SEP-2000; 2000US-235840P.
27-SEP-2000; 2000US-235863P.
28-SEP-2000; 2000US-236028P.
28-SEP-2000; 2000US-236032P.
28-SEP-2000; 2000US-236033P.
28-SEP-2000; 2000US-236034P.
28-SEP-2000; 2000US-236109P.
28-SEP-2000; 2000US-236111P.
29-SEP-2000; 2000US-236842P.
29-SEP-2000; 2000US-236891P.
02-OCT-2000; 2000US-237172P.
02-OCT-2000; 2000US-237173P.
02-OCT-2000; 2000US-237278P.
02-OCT-2000; 2000US-237294P.
02-OCT-2000; 2000US-237295P.
02-OCT-2000; 2000US-237316P.
03-OCT-2000; 2000US-237425P.
03-OCT-2000; 2000US-237598P.
03-OCT-2000; 2000US-237604P.
03-OCT-2000; 2000US-237606P.
03-OCT-2000; 2000US-237608P.
01-NOV-2000; 2000US-244867P.
01-NOV-2000; 2000US-245084P.
XX
(VAL-) AVALON PHARM.
XX
YOUNG PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
Soppet DR, Weaver Z;
WPI; 2002-188264/24.
XX
Screening for anti-neoplastic agent involves exposing cells to a
chemical agent to be tested for anti-neoplastic activity, and
determining a change in expression of a gene of a signature gene set
-
Claim 1; SEQ ID 6875; 44pp; English.
XX
The present invention describes a method (M1) for screening for an
anti-neoplastic agent. The method involves exposing cells to a chemical
agent to be tested for anti-neoplastic activity, determining a change in
expression of at least one gene (I) of a signature gene set, where (I)
comprises a sequence (S) selected from 8447 sequences (given in ABL61664
to ABL70110), or is at least 95% identical to (S), where a change in
expression is indicative of anti-neoplastic activity. (I) has cytostatic
activity and can be used in gene therapy. M1 can be used for screening
an anti-neoplastic agent, and can be used for producing a product which
is the data collected with respect to the anti-neoplastic agent as a
result of M1, and the data is sufficient to convey the chemical
structure and/or properties of the agent. M1 can be used in the
treatment of cancer such as colon, breast, stomach, lung, thyroid,
oesophageal, ovarian, kidney, prostate or pancreatic cancer,
adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
carcinoma, papillary carcinoma and Wilm's tumour.
XX
Sequence 14770 BP; 4829 A; 3240 C; 3211 G; 3490 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 719 Length: 14770
Score: 79.50 Matches: 28
Percent Similarity: 43.81% Conservative: 18
Best Local Similarity: 26.67% Mismatches: 52
Query Match: 11.76% Indels: 7
DB: 24 Gaps: 4
US-10-048-196-2 (1-134) x ABL68538 (1-14770)
QY 14 AlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro---LysAsnSerSer 32
DB 11890 GCCCTTACTTCTTCTTCATGCTAGATGTAAGTCCAGATTCAGTGAACACACACCC 11949
```


QY 33 AlaAsnLeuThrThrSerLeuLeuLysHisAlaValLysGlnThrCysGlnThrGlnLeu 52
 Db 11950 AGGGT-----AACATAATT-----GCAGTTAGAAAAGCATGTGCCACACAAAAG 11994
 QY 53 ThrGlyHisGlnThrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLys 72
 Db 11995 CAAGGG---CAGCCAGAGAAAGCAAGGCCAACACAGCTTCCATCCCAAGTTGCCAGTAAG 12051
 QY 73 IleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThr 92
 Db 12052 GTAAGATCCACCTGTGTCACTACCACCACCACCACCTGCCACACCACCACCTACCACC 12111
 QY 93 GluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleVal 112
 Db 12112 ACTACCACCACCACCACCTGCACAGTTAAAGTTAGGAAAAGTCAGCTCAAGGAAGTATGT 12171
 QY 113 ArgHisSerLeuLys 117
 Db 12172 AACATTCCATTGAA 12186
 RESULT 30
 ID AAS65159/c
 XX AAS65159 standard; cDNA; 14797 BP.
 AC AAS65159;
 DT 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #963.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX WO200175067-A2.
 PD 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR P-PSDB; ABG00972.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 1; SEQ ID No 963; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 14797 BP; 3493 A; 3212 C; 3266 G; 4826 T; 0 other;
 Alignment Scores: 721 Length: 14797
 Pred. No.: 79.50 Matches: 28
 Score: 43.81% Conservative: 18
 Best Similarity: 26.67% Mismatches: 52
 Query Match: 11.76% Indels: 7
 DB: 23 Gaps: 4
 US-10-048-196-2 (1-134) x AAS65159 (1-14797)
 QY 14 AlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro---LysAsnSerSer 32
 Db 2908 GCCCTTACTTCTTCTTCATGTGTAGATGTAAGTCCAGAAATTCAGTGAAAAACACACCCC 2849
 QY 33 AlaAsnLeuThrThrSerLeuLysHisAlaValLysGlnThrCysGlnThrGlnLeu 52
 Db 2848 AGGGT-----AACATAATT-----GCAGTTAGAAAAGCATGTGCCACACAAAAG 2804
 QY 53 ThrGlyHisGlnThrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLys 72
 Db 2803 CAAGGG---CAGCCAGAGAAAGCAAGGCCAACACAGCTTCCATCCCAAGTTGCCAGTAAG 2747
 QY 73 IleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThr 92
 Db 2746 GTAAGATCCACCTGTGTCACTACCACCACCACCACCTGCCACACCACCACCTACCACC 2687
 QY 93 GluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleVal 112
 Db 2686 ACTACCACCACCACCACCTGCACAGTTAAAGTTAGGAAAAGTCAGCTCAAGGAAGTATGT 2627
 QY 113 ArgHisSerLeuLys 117
 Db 2626 AACATTCCATTGAA 2612
 RESULT 31
 ID AAS86131
 XX AAS86131 standard; cDNA; 15187 BP.
 AC AAS86131;
 DT 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #21935.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX WO200175067-A2.
 PD 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR P-PSDB; ABG21944.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX PS Claim 1: SEQ ID No 21935; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 15187 BP; 4913 A; 3361 C; 3318 G; 3594 T; 1 other;

Alignment Scores:
 Pred. No.: 748 Length: 15187
 Score: 79.50 Matches: 28
 Percent Similarity: 43.81% Conservative: 18
 Best Local Similarity: 26.67% Mismatches: 52
 Query Match: 11.76% Indels: 7
 DB: 23 Gaps: 4

US-10-048-196-2 (1-134) x AAS86131 (1-15187)

QY 14 AlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro---LysAsnSerSer 32
 DB 11904 GCCTTACTACTCTTCATGTGTAGATGTAAGTCCAGATTCCAGTGAACACACAC 11963
 QY 33 AlaAsnLeuThrThrSerLeuLeuLysHisAlaValLysGlnThrCysGlnThrGlnLeu 52
 DB 11964 AGGGAT-----ACATAATT-----GCAGTTAGAAAGCATGTGCCACACAAAG 12008
 QY 53 ThrGlyHisGlnTyrTrpLysIleAlaAlaMetLysLeuSerSerGluSerLysAlaLys 72
 DB 12009 CAAGGG---CAGCCAGAGAAGCAAGGCAAGCAACAGCTTCCATCCCAAGTTGCCAGTAAAG 12065
 QY 73 IleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeuThr 92
 DB 12066 GTAAGATCCACCTGTCTACTACCCACACACACCTGCCACACACACCTACCTACCC 12125
 QY 93 GluLeuThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGlnLysIleVal 112
 DB 12126 ACTACACACACACACAGCTGCACAGTAAAGTTAGGAAAGTCAGCTCAAGGAAGTATGT 12185
 QY 113 ArgHisSerLeuLys 117
 DB 12186 AAACATTCCATTGAA 12200

RESULT 32

AAA81479

ID AAA81479 standard; DNA; 69936 BP.

AC AAA81479;

XX 04-DEC-2000 (first entry)

XX DE N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27. *
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX OS Neisseria meningitidis.
 XX PN WO2000022430-A2.
 XX PD 20-APR-2000.
 XX PF 08-OCT-1999; 99WO-US235573.
 XX PR 09-OCT-1998; 98US-0103794.
 XX PR 30-APR-1999; 99US-0132068.
 XX PA (CHIR) CHIRON CORP.
 XX PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 DR WPI; 2000-318079/27.
 XX PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea
 XX PS Claim 7: Page 547-567; 1760pp; English.

XX CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX SQ Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

Alignment Scores:

Pred. No.: 6.15e+03 Length: 69936
 Score: 79.50 Matches: 26
 Percent Similarity: 40.57% Conservative: 17
 Best Local Similarity: 24.53% Mismatches: 46
 Query Match: 11.76% Indels: 17
 DB: 21 Gaps: 4

US-10-048-196-2 (1-134) x AAA81479 (1-69936)

QY 24 ProGluSerAsnProLysAsnSerSerAlaAsnLeu---ThrThrSerLeuLysHis 42

DB 65504 CCGAAGCGGACGACCTACCGGACCACTGGGCAACGCGCATCAAGGCA 65563

(CORI-) CORIXA CORP.
Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
L'maisonneuve J, Zhang Y, Jen S, Carter D;
WPI: 2001-616774/71.
Propionibacterium acnes polypeptides and nucleic acids useful for
vaccinating against and diagnosing infections, especially useful for
treating acne vulgaris -
Claim 1: SEQ ID No 161: 1069pp; English.
Sequences AAS59506-AAS59804 represent DNA molecules encoding
Propionibacterium acnes immunogenic polypeptides. The proteins and their
associated DNA sequences are used in the treatment, prevention and
diagnosis of medical conditions caused by P. acnes. The disorders include
SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and
osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
in infections of bone, joints and the central nervous system, however it
is particularly involved in the inflammatory lesions associated with acne
vulgaris. A method for detecting the presence or absence of P. acnes in a
patient comprises contacting a sample with a binding agent that binds to
the proteins of the invention and determining the amount of bound protein
in the sample. The polypeptides may be used as antigens in the production
of antibodies specific for P. acnes proteins. These antibodies can be
used to downregulate expression and activity of P. acnes polypeptides and
therefore treat P. acnes infections. The antibodies may also be used as
diagnostic agents for determining P. acnes presence, for example, by
enzyme linked immunosorbent assay (ELISA). This sequence encodes the
polypeptides shown in AAU65346-AAU65365 and AAU67770.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 2243 BP: 435 A; 627 C; 753 G; 428 T; 0 other;

Alignment Scores:			
Pred. No.:	60.8	Length:	2243
Score:	79.00	Matches:	39
Percent Similarity:	38.51%	Conservative:	18
Best Local Similarity:	26.35%	Mismatches:	61
Query Match:	11.69%	Indels:	30
DB:	23	Gaps:	6
US-10-048-196-2 (1-134) x AAS59666 (1-2243)			
Qy	9	AlaThrLeuMetThrAlaPheThrLeuAlaSerCysAlaSerThrProGluSerAsnPro	28
	:::		
Db	907	AGTACAACGTCAGCGCGGCTCTTCGCGGCTCTTCGCAACACGCGCGCTATTACCA	848
Qy	29	LysAsnSerSerAlaAsnLeu-----ThrThrSerLeuLeuLysHisAla	43
	::: :::		
Db	847	GCCTGCTGACGATCCGGTTGGACATCTTGTTCGGTCACCTCATCTGCTGGTGAGTCG	788
Qy	44	VallLysGlnThr-----CysGlnThrGlnLeuThrGlyHisGlnTrpTrpLysIle	60
	:::		
Db	787	ACGACCATTATCACCCTGGGTGCCCTCTCACGGAGTCGCCCCACAG-----	740
Qy	61	AlaAlaMetLysLeuSerSerGluSerLysAlaLysIleSerGluThrAlaCysGly----	79
	:::		
Db	739	TGCACATGGACTCAACTTCACCGAGCGTCTCGGTTCGGTGTCCGCGCGCGCTCGGGGTCT	680
Qy	80	-----CysValAlaAsnLysAlaProGluAlaVal	89
Db	679	CCGACCCTGGCCATAGTGTGCCAGCCAGCTGCGGTCCGCGCAAGACATCGACGTCGACCGTC	620
Qy	90	SerLeuThrGluLeuThrThrAlaAlaIleAsnProAsnAlaArgThrGluValAlaGln	109
Db	619	TCVTGGACAACTGGGAGTCGGATGGCTACACACGGAGGCACTCAGCGCGCTTGTCTAC	560
Qy	110	LysIleValArqHisSerLeuLysPro-----CysMetLeuGluThrValAsn	125

```

Db      559 -----||| ::::::::::||| ||| ||:::
          AGACGCGGTCTAGACCGCAAGGCTGTCATGGCAGCTGCCACGCCTAACG 509

Qy      126 AlaPheIleValProThrThrThr 133
          ||| ::: |||||||:::
Db      508 GCACCGTTATCTCTACGACCTCA 485

RESULT 35
AAS68171/c
ID   AAS68171 standard; cDNA; 3300 BP.
XX     AC
XX     AAS68171;
XX     XX
XX     XX
DT     DT
DE     DE
DE     DNA encoding novel human diagnostic protein #3975.
XX     XX
KW     Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW     food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX     OS
XX     Homo sapiens.
XX     WO200175067-A2.
XX     PN
XX     PD
XX     11-OCT-2001.
XX     PF
XX     30-MAR-2001; 2001WO-US08631.
XX     PR
XX     31-MAR-2000; 2000US-0540217.
XX     PR
XX     23-AUG-2000; 2000US-0649167.
XX     PA
XX     (HYSE-) HYSEQ INC.
XX     PI
XX     Drmanac.RT, Liu C, Tang YT;
XX     WPI; 2001-639362/73.
XX     DR
XX     P-PSDB; ABG03984.
XX     PT
XX     New isolated polynucleotide and encoded polypeptides, useful in
XX     diagnostics, forensics, gene mapping, identification of mutations
XX     responsible for genetic disorders or other traits and to assess
XX     biodiversity -
XX     Claim 1; SEQ ID NO 3975; 103pp; English.

```

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of and products dependent on DNA and amino acid sequences. AA564157-AA594564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct_sequences.

Sequence 3300 BP: 1184 A: 745 C: 669 G: 702 T: 0 other;
XX

Alignment Scores:		
Pred. No.:	104	Length: 3300
Score:	79.00	Matches: 36
Percent Similarity:	40.74%	Conservative: 19


```
CC at.ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3300 BP; 1184 A; 745 C; 669 G; 702 T; 0 other;

Alignment Scores:
Pred. No.: 104 Length: 3300
Score: 79.00 Matches: 36
Percent Similarity: 40.74% Conservative: 19
Best Local Similarity: 26.67% Mismatches: 42
Query Match: 11.69% Indels: 38
DB: 23 Gaps: 7

US-10-048-196-2 (1-134) x AAS73830 (1-3300)
QY 5 LeuTyrValThrAlaThrLeuMetThrAlaPhe-----ThrLeuAlaSerCysAlaSer 22
Db 2591 CTTTCCCTCACATCAACATGATGGCTCCATCAGAGTAACCTTTGCAGGAGTTGCATGC 2532
QY 23 ThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle----- 40
Db 2531 TTAGGGCAAAGAGGCGCTATATTAGTTCTCCCAATTCTTCCCTCCACTCTCTTGGGTATC 2472
QY 41 -----LysHisAla-----Val 44
Db 2471 CTGACATTCCTGGGCCCATGCTCCAAAGCCTTCTCTGGTTGTTGTGCAGACTGCTCGGGAGCC 2412
QY 45 LysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr-----TrpLys 59
Db 2411 AAACAATCCAAAGCCACTTTTGTGCATCGCCATGAGCAATCACTGCTCAAGGACTGGAAA 2352
QY 60 ---IleAlaAlaMetLysLeuSer-----SerGluSerLysAla 71
Db 2351 CACGTAGAATCCATGCATCTATTCGTATTCTTGTTCAGTGGGAGATCCAGATGGGAATGAG 2292
QY 72 LysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeu 91
Db 2291 AGGTGTCTGAGACAGCTCTCTTCTCTCATACATTTAGCATATTTAGCATATGCTCTCATG 2232
QY 92 -----ThrGluLeuThrThrAlaAlaIleAsnPro 101
Db 2231 GTAATAGTGCCTTGTCTGTAAGAAAGAACTTCAACCCCAATCCA 2187

RESULT 39
AAS74991/C
ID AAS74991 standard; cDNA; 3300 BP.
XX
AC AAS74991;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #10795.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT.
XX
DR WPI: 2001-639362/73.
DR P-PSDB; ABG10804.
XX
```

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity

Claim 1: SEQ ID No.10795; 103pp: English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 3300 BP; 1184 A; 745 C; 669 G; 702 T; 0 other;

Alignment Scores:
Pred. No.: 104 Length: 3300
Score: 79.00 Matches: 36
Percent Similarity: 40.74% Conservative: 19
Best Local Similarity: 26.67% Mismatches: 42
Query Match: 11.69% Indels: 38
DB: 23 Gaps: 7

US-10-048-196-2 (1-134) x AAS74991 (1-3300)

QY 5 LeuTyrValThrAlaThrLeuMetThrAlaPhe-----ThrLeuAlaSerCysAlaSer 22
Db 2591 CTTTCCCTCACATCAACATGATGGCTCCATCAGAGTAACCTTTGCAGGAGTTGCATGC 2532

QY 23 ThrProGluSerAsnProLysAsnSerSerAlaAsnLeuThrThrSerLeuIle----- 40
Db 2531 TTAGGGCAAAGAGGCGCTATATTAGTTCTCCCAATTCTTCCCTCCACTCTCTTGGGTATC 2472

QY 41 -----LysHisAla-----Val 44
Db 2471 CTGACATTCCTGGGCCCATGCTCCAAAGCCTTCTCTGGTTGTTGTGCAGACTGCTCGGGAGCC 2412

QY 45 LysGlnThrCysGlnThrGlnLeuThrGlyHisGlnTyr-----TrpLys 59
Db 2411 AAACAATCCAAAGCCACTTTTGTGCATCGCCATGAGCAATCACTGCTCAAGGACTGGAAA 2352

QY 60 ---IleAlaAlaMetLysLeuSer-----SerGluSerLysAla 71
Db 2351 CACGTAGAATCCATGCATCTATTCGTATTCTTGTTCAGTGGGAGATCCAGATGGGAATGAG 2292

QY 72 LysIleSerGluThrAlaCysGlyCysValAlaAspLysAlaProGluAlaValSerLeu 91
Db 2291 AGGTGTCTGAGACAGCTCTCTTCTCTCATACATTTAGCATATTTAGCATATGCTCTCATG 2232

QY 92 -----ThrGluLeuThrThrAlaAlaIleAsnPro 101
Db 2231 GTAATAGTGCCTTGTCTGTAAGAAAGAACTTCAACCCCAATCCA 2187

RESULT 40
AAS81666/C
ID AAS81666 standard; cDNA; 3300 BP.

This Page Blank (uspto)